Rec'd PCT/PTO 31 JAN 2005

REPUBLIKA SLOVENIJA
istrstvo za gospodarstvo

T/EP 03/08308
10/522827

URAD REPUBLIKE SLOVENIJE ZA INTELEKTUALNO LASTNINO

Potrdilo

REC'D 08 SEP 2003
WIPO PCT

Certificate

Urad Republike Slovenije za intelektualno lastnino potrjuje, da je priloženi dokument istoveten z izvirnikom patentne prijave, kot sledi:

Slovenian Intellectual Property Office hereby certifies that the document annexed hereto is a true copy of the patent application, as follows:

(22) Datum prijave (*Application Date*):

31.7.2002 (31.jul.2002)

(21) Številka prijave (*Application No.*):

P-200200188

(54) Naziv (*Title*):

Sintetski gen za humani granulocitne kolonije stimulirajoči dejavnik za ekspresijo v E. coli

Ljubljana, 3.6.2003

Janez Kukec-Mezek
svetovalec Vlade

REPUBLIKA SLOVENIJA
URAD RS ZA INTELEKTUALNO LASTNINO

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

REPUBLIKA SLOVENIJA
MINISTRSTVO ZA GOSPODARSTVO

URAD REPUBLIKE SLOVENIJE
ZA INTELEKTUALNO LASTNINO
1000 LJUBLJANA, KOTNIKOVA 6

ZAHTEVA ZA PODELITEV PATENTA

1. Naslov za obveščanje:

Lek farmacevtska družba d. d.

Sektor intelektualna lastnina

Verovškova 57

1526 Ljubljana

Slovenija

tel.:01 580 2005
faks: 01 568 2123

šifra:bk/910

Potrdilo o prejemu prijave
(izpolni urad)

Datum vložitve prijave: 31.7.2002

Številka prijave: P-2002-00188

Žig urada in podpis:

2. Prijavitelj (priimek, ime in naslov, za pravne osebe firma in sede):
Lek farmacevtska družba d. d.
Verovškova 57
1526 Ljubljana
Slovenija

Registrska številka:

3. Zastopnik:

4. Izumitelj (priimek, ime in naslov):
Jevševar Simona
Vodole 26
2000 Maribor

5. Naziv izuma:
Sintetski gen za humani granulocitne kolonije stimulirajoči dejavnik za ekspresijo v *E. coli*

6. Podatki o zahtevani prednostni pravici in podlagi zanjo:

7. Dodatne zahteve:
- ☐ prijava je za patent s skrajšanim trajanjem
- ☐ predhodna objava patenta po preteku _____ mesecev
- ☐ prijava je izločena iz prijave številka: __________

8. Izjava:
- ☐ izjava o skupnem predstavniku:

REPUBLIKA SLOVENIJA
MINISTRSTVO ZA GOSPODARSTVO
URAD RS ZA INTELEKTUALNO LASTNINO

Prejeto dne: 31-07-2002		Osebna oddaja: ☒
Podpis:	Oddano priporočeno dne:	
Šifra:	Poštna številka: 12841	

REPUBLIKA SLOVENIJA
MINISTRSTVO ZA GOSPODARSTVO

URAD REPUBLIKE SLOVENIJE
ZA INTELEKTUALNO LASTNINO
1000 LJUBLJANA, KOTNIKOVA 6

9. Priloge:

- [x] opis izuma, ki ima __17__ strani
- [x] patentni zahtevek (zahtevki), ki ima(jo) __1__ strani; število zahtevkov: _13___
- [x] skice (če so zaradi opisa izuma potrebne); število listov: __4__
- [x] povzetek
- [ ] potrdilo o plačilu prijavne pristojbine
- [ ] potrdilo o deponiranju biološkega materiala, če gre za izum, ki ga ni mogoče drugače opisati
- [ ] pooblastilo zastopniku
- [ ] generalno pooblastilo zastopniku je deponirano pri uradu pod št.: ________
- [ ] potrdilo o razstavni prednostni pravici
- [ ] podatki o drugih prijaviteljih
- [x] podatki o drugih izumiteljih
- [ ] prikaz zaporedja nukleotidov ali aminokislin v opisu
- [x] prijava je bila predhodno posredovana po faksu ali v elektronski obliki
- [x] _Izjava prijavitelja o prikazu zaporedja____

___Alenka Košak____________________
Priimek in ime ter podpis prijavitelja (zastopnika)

Podatki o drugih izumiteljih:

Viktor Menart
Logatec, Dol 3
1370 Logatec

Naziv izuma

Sintetski gen za humani granulocitne kolonije stimulirajoči dejavnik za ekspresijo v *E. coli*

Področje tehnike

Predloženi izum se nanaša na sintetski gen za humani granulocitne kolonije stimulirajoči dejavnik (hG-CSF), ki omogoča ekspresijo v *E. coli* z nivojem ekspresije, ki je enak ali višji od 52% rekombinantnega hG-CSF glede na celotne proteine po ekspresiji.

hG-CSF uvrščamo med kolonije stimulirajoče dejavnike, ki regulirajo diferenciacijo in proliferacijo hematopoetskih celic sesalcev. Imajo odločilno vlogo pri tvorbi nevtrofilcev, zato so primerni za uporabo v medicini na področju hematologije in onkologije.

Za klinično uporabo sta danes na trgu dve obliki: lenograstim, ki je glikoziliran, in ga pridobivajo z ekspresijo v sesalski celični liniji, ter filgrastim, ki je neglikoziliran, in ga pridobivajo z ekspresijo v bakteriji *Escherichia coli* (*E. coli*).

Bistvo predloženega izuma

Bistvo predloženega izuma je, da je mogoče s sintetskim genom, ki kodira za hG-CSF, doseči nivo ekspresije (akumulacijo) v *E. coli*, ki je enak ali višji od 52% rekombinantnega hG-CSF glede na celotne proteine v *E. coli*. Pri ekspresiji se uporabi ekspresijski plazmid z močnim T7 promotorjem. Sintetski gen za hG-CSF se pripravi s kompleksno kombinacijo dveh postopkov, ki omogočata pripravo optimiziranega sintetskega gena za hG-CSF za ekspresijo v *E. coli*. Prvi postopek je ta, da se na nekaterih mestih zamenja neugodne kodone za ekspresijo v *E. coli* z bolj ugodnimi kodoni za *E. coli*. Drugi pa, da se na nekaterih mestih zamenja GC bogate regije z AT bogatimi regijami. Na nekaterih mestih se sintetski gen za G-CSF, ki je predmet izuma, pripravi tako, da se uporabi ena od obeh metod, na nekaterih mestih kombinacija obeh navedenih metod, nekatera mesta se pa ne spremenijo. Pri pripravi sintetskega gena za hG-CSF, ki je tudi predmet izuma, ne spreminjamo regij, ki so izven kodirajočega področja. Tako ne uvajamo sprememb v področju iniciacije

translacije (TIR), v področju vezavnega mesta za ribosom (RBS) in v področju razdalje med start kodonom in RBS.

Stanje tehnike

Vpliv več zaporednih neugodnih kodonov, kot so argininiski (AGG/AGA; CGA), leucinski (CTA), izoleucinski (ATA) ali prolinski (CCC), na uspešnost translacije in s tem zmanjšanje količine ali kvalitete nastalega proteina izraženega v *E. coli*, je opisan v Kane JF, Current Opinion in Biotechnology, 6:494-500 (1995). Podoben je vpliv posameznih neugodnih kodonov, če se pojavljajo na različnih mestih.

Vpliv na uspešnost translacije v *E. coli* imajo tudi GC bogate regije, če pride zaradi njih pri sekundarni strukturi mRNA do nastanka stabilne dvoverižne RNA. Vpliv je največji, kadar so GC bogate regije mRNA na mestu, kjer se veže ribosom, v neposredni bližini vezave ribosoma ali v neposredni bližini start kodona (Makrides SC, Microbiological Reviews, 60:512-538 (1996); Baneyx F, Current Opinion in Biotechnology, 10:411-421 (1999)). Znanih je več metod ocenjevanja sekundarne strukture in izračunavanja minimalne proste energije posamezne RNA molekule, kar naj bi bilo osnovno merilo za najbolj stabilno oziroma najbolj verjetno strukturo (SantaLucia J Jr in Turner DH, Biopolymers, 44:309-319 (1997)). Zanesljivi algoritmi za napoved prave sekundarne strukture, razen v nekaterih primerih še niso poznani; prav tako še ni možno dokazati kvantitativne relacije z nivojem ekspresije (Smit MH in van Duin JJ. Mol. Biol., 244, 144-150 (1994)). Trodimenzionalnih struktur proteinov tudi še ni možno predvideti (Tinoco I in Bustamante C, J. Mol. biol, 293:271-281 (1999)).

Povečanje nivoja ekspresije z optimizacijo DNA zaporedja v TIR, RBS in razdalje med start kodonom in mestom RBS je opisano v McCarthy JEG in Brimacombe R, Trends Genet 10:402-407 (1994). Vzrok za povečanje nivoja ekspresije v tem primeru je bolj učinkovit pričetek translacije in tekočega nadaljevanja v kodirajoče področje mRNA.

Pridobitev dovolj velikih količin hG-CSF za *in vitro* biološke študije z ekspresijo v *E. coli* je opisana v Souza LM et al, Science 232:61-65 (1986) in v Zsebo KM et al, Immunobiology 172:175-184 (1986). Dosežen nivo ekspresije hG-CSF je bil manjši od 1%.

V patentu US4810643 je opisana uporaba sintetskega gena za hG-CSF, ki je bil skonstruiran predvsem na osnovi zamenjave neugodnih kodonov s kodoni, ki so za *E. coli* optimalni. V kombinaciji s termoinducibilnim promotorjem iz lambda faga je bil dosežen nivo ekspresije hG-CSF od 3 do 5% glede na celotne celične proteine, kar ne omogoča ekonomične proizvodnje v industrijskem merilu.

8-10% akumulacijo hG-CSF glede na celotne celične proteine so dosegli s spremembo prvih štirih kodonov v 5' področju gena za hG-CSF, kot je opisano v Wingfield P et al, Biochem. J, 256:213-218 (1988).

Ekspresija hG-CSF v *E. coli* z izkoristkom do 17% hG-CSF glede na celotne bakterijske proteine je opisana v Devlin PE et al, Gene 65:13-22 (1988). Tak izkoristek so dosegli z delno optimizacijo DNA zaporedja na 5' koncu gena (kodoni za prve štiri aminokisline), pri čemer so spremenili GC v AT bogato regijo in z uporabo relativno močnega promotorja iz faga lambda in optimizacije DNA zaporedja na 5' koncu gena (kodoni za prve štiri aminokisline). Nivo ekspresije ni zelo visok, kar prispeva k slabšim izkoristkom pridobivanja in manjšo ekonomičnost v industrijskem merilu.

Uporaba sintetskega gena in okoli 30% ekspresija sta opisani v Kang SH et al, Biotechnology letters, 17(7):687-692 (1995). Ta nivo so dosegli z uvedbo za *E. coli* ugodnih kodonov, z modifikacijami v TIR in z dodatnimi modifikacijami setov kodonov, pri tem da niso bistveno spreminjali 3' konca gena. Za dosego navedenega nivoja ekspresije so bile potrebne spremembe gena za hG-CSF tudi v TIR, pri tem da nivo ekspresije ni presegel 30%.

V patentu US5840543 je opisan sintetski gen za hG-CSF, ki je bil skonstruiran z uvajanjem AT bogatih regij na 5' konec gena in z zamenjavo neugodnih kodonov s kodoni, ugodnimi za *E. coli*. Pod kontrolo Trp promotorja so dosegli ekspresijo z izkoristkom 11% hG-CSF glede na celotne celične proteine. Z dodatkom leucina in treonina ali njune kombinacije v fermentacijski medij, v katerem so bakterije gojene, pa so dosegli do 35% akumulacijo hG-CSF glede na celotne celične proteine. Nivo ekspresije so v tem primeru povečali z dodajanjem aminokislin v fermentacijski medij, kar predstavlja dodatni strošek v postopku pridobivanja hG-CSF in ni ekonomično v industrijskem merilu. Zgolj optimizacija gena za hG-CSF pa ni omogočala večjega nivoja ekpresije hG-CSF.

Do sedaj najvišja akumulacija hG-CSF glede na celotne celične proteine je opisana v Jeong et al, Protein Expression and Purification 23,:311-318 (2001) in je 48%. Dosegli so jo z spremembno N-terminalnega dela gena in z indukcijo z 1 mM IPTG.

V splošnem ne obstajajo zapisi, da bi bilo mogoče predvideti, kakšen bo nivo ekspresije nativnih humanih genov v prokariotskih organizmih, kot je npr. bakterija *E. coli*. Opisani nivoji ekspresije so relativno nizki ali celo komaj zaznavni tudi pri uporabi ekspresijskih plazmidov z močnimi promotorji, kot je npr. promotor iz lambda ali T7 faga. Iz literature sledi, da na visoko akumulacijo humanega proteina v *E. coli* vpliva mnogo parametrov (neugodni kodoni ali njihovo klastriranje; področja z velikim številom GC baznih parov, neugodne sekundarne strukture mRNA, nestabilna mRNA).

Zaenkrat ni nobenega povsem izdelanega pravila, kako kombinirati kodone, da bo imela rezultirajoča mRNA za ekspresijo optimalno sekundarno in terciarno strukturo. Matematični in strukturni modeli za napoved in termodinamsko stabilnost sekundarnih struktur so sicer dostopni, vendar še preveč nezanesljivi že pri sekundarnih strukturah, pri terciarnih pa jih sploh še ni. Ti trenutno izdelani modeli torej ne omogočajo napovedi, kakšen je vpliv različnih kombinacij kodonov na ekspresijo.

Ne v patentni ne v strokovni literaturi ni opisanih učinkovitejših načinov za rešitev problema nizkega nivoja ekspresije nativnega gena za hG-CSF v *E. coli*.

Opis izuma

Ugotovili smo, da lahko problem nizkega nivoja ekspresije nativnega gena za hG-CSF v *E. coli* rešimo s pomočjo optimizacije nativnega gena za hG-CSF in s tem pripravo sintetskega gena za G-CSF. V primerjavi z do sedaj znanimi podatki na ta način presenetljivo dobimo bistveno višji nivo ekspresije.

Z izrazom 'hG-CSF' je mišljen humani granulocitne kolonije stimulirajoči dejavnik, ki zajema tudi rekombinantni hG-CSF, ki ga dobimo z ekspresijo v *E. coli*.

Sintetski gen za hG-CSF, ki je predmet izuma, smo pripravili tako, da smo spremenili nukleotidno zaporedje gena za nativni hG-CSF, pri čemer se aminokislinsko zaporedje ni spremenilo in je ostalo identično nativnemu hG-CSF.

Predloženi izum nadalje zajema ekspresijo na tak način pridobljenega sintetskega gena v *E. coli* in nivo ekspresije na tak način pridobljenega sintetskega gena.

Z izrazom 'nivo ekspresije' je mišljen delež hG-CSF, ki ga dobimo po heterologni ekspresiji gena za hG-CSF v *E. coli*, glede na celotne proteine, ki so prisotni po ekspresiji.

Z izrazom 'heterologna ekspresija' je mišljena ekspresija tistih genov, ki so tuji organizmu, v katerem poteka ekspresija.

Z izrazom 'homologna ekspresija' je mišljena ekspresija tistih genov, ki so lastni organizmu, v katerem poteka ekspresija.

Z izrazom 'ugodni kodoni' so mišljeni tisti kodoni, ki jih določen organizem (npr. *E. coli*) uporablja za produkcijo največ mRNA molekul. Te kodone organizem uporablja za gene z visoko homologno ekspresijo.

Z izrazom 'neugodni kodoni' so mišljeni tisti kodoni, ki jih določen organizem (npr. *E. coli*) uporablja le za gene z nizko ekspresijo. Ti kodoni so v določenem organizmu redko uporabljeni (nizka homologna ekspresija).

Z izrazom 'GC bogate regije' so mišljena tista področja v genu, v katerih prevladujeta bazi gvanin (G) in citozin (C).

Z izrazom 'AT bogate regije' so mišljena tista področja v genu, v katerih prevladujeta bazi adenin (A) in timin (T).

Z izrazom 'sintetski gen' je mišljen gen, ki se od nativnega razlikuje po DNA zaporedju, aminokislinsko zaporedje pa ostane nespremenjeno. Pridobljen je s tehnikami rekombinantne tehnologije DNA.

Z izrazom 'nativni gen' je mišljen gen, ki ni spremenjen s tehnikami rekombinantne DNA tehnologije.

Z izrazom 'segment' so mišljeni posamezni deli gena, ki so na obeh straneh omejeni z enojnimi restrikcijskimi mesti, katera služijo za subkloniranje sintetsko pripravljenih delov gena.

Z izrazom 'segment I' je mišljen 5' konec gena za hG-CSF med restrikcijskima mestoma Nde I (3) in Sac I (194), t.j. 191 bp dolgo zaporedje, ki je bilo sintetizirano na novo.

Z izrazom 'segment II' je mišljen del gena za hG-CSF med restrikcijskima mestoma Sac I (194) in Apa I (309), t.j. 115 bp dolg osrednji dela gena, ki je bil sintetiziran na novo.

Z izrazom 'segment III' je mišljen del gena za hG-CSF med restrikcijskima mestoma Apa I (309) in Nhe I (467), t.j. 158 bp dolg dela gena, kjer je z izjemo Arg148 in Gly150 ohranjeno nativno DNA zaporedje za hG-CSF.

Z izrazom 'segment IV' je mišljen 3' konec gena za hG-CSF med restrikcijskima mestoma Nhe I (467) in BamH I (536), t.j. 69 bp dolg končni del gena, ki je bil sintetiziran na novo.

Sintetski gen za hG-CSF, ki je predmet izuma, je pripravljen s kombinacijo naslednjih metod:

- zamenjavo za *E. coli* neugodnih kodonov s kodoni, ki so za *E. coli* ugodni: v segmentu II (med restrikcijskima mestoma Sac I (194) in Apa I (309)) in segmentu IV (med restrikcijskima mestoma Nhe I (467) in BamH I(536))
- zamenjavo nekaterih GC bogatih regij z AT bogatimi regijami, s tem, da odpravimo tudi za *E. coli* najbolj neugodne kodone, pri tem pa večinoma ne uporabimo za *E. coli* najbolj optimalnih kodonov: v segmentu I (med restrikcijskima mestoma Nde I (3) in Sac I (194)).
- popolnoma nespremenjenega nativnega zaporedja 46 kodonov (med Pro102 in Arg147) znotraj segmenta III.
- odpravo dveh za *E. coli* neugodnih kodonov (Arg148 in Gly150) na koncu segmenta III.

Optimizacija gena za hG-CSF, ki je predmet izuma, ne vključuje sprememb področij TIR, področij RBS ter področij v razdalji med start kodonom in RBS.

Sintetski gen za hG-CSF, ki je predmet izuma, omogoča, da se ob ekpresiji pridobljenega sintetskega gena za hG-CSF doseže nivo ekspresije hG-CSF v *E. coli*, ki je enak ali višji od 52%, omogočeno je tudi doseganje nivoja ekspresije okoli 55% ali celo okoli 60%. Visok nivo ekspresije sintetskega gena za hG-CSF, ki je predmet izuma, omogoča visoke izkoristke pri pridobivanju hG-CSF, hitrejšo in bolj enostavno čiščenje in izolacijo heterolognega hG-CSF, lažjo medprocesno kontrolo, boljšo ekonomičnost celotnega procesa in s tem omogoča učinkovito pridobivanje hG-CSF

v industrijskem merilu. Tako pridobljen hG-CSF je primeren za klinično uporabo v medicini.

Priprava sintetskega gena za hG-CSF, ki je predmet izuma, se prične s predpripravo nativnega gena za hG-CSF in plazmidov. Gen za nativni hG-CSF je lahko humanega izvora, prav tako lahko uporabimo isti princip za vse gene, ki so homologni v tistih regijah, ki vsebujejo enojna restrikcijska mesta, katera uporabimo za subkloniranje na novo sintetiziranih segmentov gena. Plazmid za izvedbo mutageneze je bil izbran tako, da je omogočal zaporedno uvajanje točkovnih mutacij. Selekcija oziroma bistvena obogatitev plazmidov z želeno mutacijo je bila dosežena s hkratno spremembo restrikcijskega mesta, in sicer iz EcoRI v EcoRV ali obratno (Transformer™ Site-Directed Mutagenesis Kit (Clontech)). Gen in plazmid se pripravita tako, da je omogočeno tudi uvajanje mutacij s kasetno mutagenezo.

Po predpripravi nativnega gena za hG-CSF in plazmidov se izvede optimizacija nativnega gena za hG-CSF in s tem priprava sintetskega gena za hG-CSF. Optimizacija se prične tako, da se gen za nativni hG-CSF razdeli v štiri (I, II, III in IV) segmente, ki so ali bodo po izvedbi oligonukleotidne mutageneze ločeni z enojnimi restrikcijskimi mesti, in se v posameznih segmentih izvede spremembe. V posameznih segmentih se izvede spremembe zaporedja gena, v določenih segmentih se pa gena ne spreminja (Slika 1). Prednostno je tako pridobljen končni optimiran sintetski gen za hG-CSF sestavljen iz delno ohranjenega nativnega zaporedja (segment III) ter 5' in 3' kodirajočih področij, ki so sintetizirana na novo (segmenti I, II in IV).

Spremembe po posameznih segmentih:

Segment I: Zamenjava za *E. coli* neugodnih kodonov z *E. coli* ugodnimi kodoni ter zamenjava GC bogatih regij z AT bogatimi regijami

Thr2 (ACC→AC**A**), Pro3 (CCC→CC**A**), Gly5 (GGC→GG**T**) Pro6 (CCT→CC**A**), Ala7 (GCC→GC**T**), Ser8 (AGC→**TC**T), Ser9 (TCC→TC**T**), Pro11 (CCC→CC**G**), Gln12 (CAG→CA**A**), Phe14 (TTC→TT**T**), Leu16 (CTC→**T**T**G**), Lys17 (AAG →AA**A**), Cys18 (TGC→TG**T**), Glu20 (GAG→GA**A**), Val22 (GTG →GT**T**), Arg23 (AGG→**C**G**T**), Lys24 (AAG→AA**A**) Ile25 (ATC→AT**T**), Gln26 (CAG→CA**A**), Gly27 (GGC→GG**T**), Gly29 (GGC→GG**T**), Ala31 (GCG→GC**T**), Leu32 (CTC→**T**T**A**), Gln33 (CAG→CA**A**), Glu34 (GAG→GA**A**), Lys35 (AAG →AA**A**), Ala38 (GCC→GC**A**), Thr39 (ACC→AC**T**), Tyr40

(TAC→TAT), Lys41 (AAG→AAA), Cys43 (TGC→TGT), His44 (CAC→CAT), Pro45 (CCC→CCA), Glu46 (GAG→GAA), Glu47 (GAG→GAA), Val49 (GTG→GTT), Leu51 (CTC→TTA), Gly52 (GGA→GGT), His53 (CAC→CAT), Gly56 (GGC→GGT), Ile57 (ATC→ATT), Pro58 (CCC→CCG), Pro61 (CCC→CCT)

Segment II: Zamenjava za *E. coli* neugodnih kodonov z *E. coli* ugodnimi kodoni
Cys65 (TGC→TGT), Pro66(CCC→CCG), Ala69 (GCC→GCG), Leu76 (TTG→CTG), Leu79 (CTC→CTG), Gly82 (GGC→GGT), Leu83 (CTT→CTG), Phe84 (TTC→TTT), Leu85 (CTC→CTG), Tyr86 (TAC→TAT), Gly88 (GGG→GGT), Leu89 (CTC→CTG), Ala92 (GCC→GCG), Gly95 (GGG→GGC), Ile96 (ATA→ATT), Pro98 (CCC→CCG), Glu99 (GAG→GAA), Leu100 (TTG→CTG), Gly101 (GGT→GGG)

Segment III: Zamenjava dveh za *E. coli* neugodnih kodonov tik pred restrikcijskim mestom NheI
Arg 148 (CGG →CGT), Gly150 (GGA→GGT)

Segment IV: Zamenjava dolgega klastra za *E. coli* neugodnih kodonov ob koncu gena z za *E. coli* ugodnimi kodoni
Gln159 (CAG→CAA), Ser160 (AGC→TCT), Phe161 (TTC→TTT), Glu163 (GAG→GAA), Val164 (GTG→GTT), Ser165 (TCG→AGC), Tyr166 (TAC→TAT), Arg167 (CGC→CGT), Leu169 (CTA→CTG), Arg170 (CGC→CGT), His171 (CAC→CAT), Leu172 (CTT→CTG), Ala173 (GCG→GCT), Pro175 (CCC→CCG)

Po pripravi sintetskega gena za hG-CSF se optimiran sintetski gen subklonira v končni plazmidni vektor, ki je izbran iz skupine pET vektorjev (Novagen), ki vsebujejo močan T7 promotor. Prednostno se uporabi plazmidni vektor pET3a. Ekspresijski plazmid, ki ob tem nastane, se nato transformira v produkcijski sev, ki je izbran iz skupine sevov, ki nosijo v kromosomu zapis za T7 RNA polimerazo, prednostno v *E. coli* BL21 (DE3).

Postopek se nadaljuje s pripravo inokuluma in fermentacijo. Fermentacijo se izvaja pri 37^0C, prednostno pri 25^0C.

Akumuliran heterologni hG-CSF se izloča v inkluzijskih telescih in je primeren za renaturacijo in za uporabo v izolacijskih postopkih.

Opis slik:

Slika 1: Shema stopenj optimizacije gena za hG-CSF

Slika 2: a) DNA zaporedje nativnega gena za hG-CSF (GenBank: NM_000759)

b) DNA zaporedje optimiranega (Fopt5) gena za hG-CSF. Poudarjeno so izpisane baze, ki se razlikujejo od nativnega gena

Slika 3: a) SDS-PAGE (4 % zgoščevalni, 15 % ločitveni; barvano s Coomassie brilliant blue) vzorcev proteinov neinducirane in inducirane kulture produkcijskih sevov *E. coli* BL21 (DE3) z ekspresijskim plazmidom pET3a pri 25° C in 42° C. Kulture so bile gojene v LBG10/amp100 mediju.

Legenda:

Nanos 1: BL21(DE3) pET3a-hG-CSF neinduciran pri 25^0C (10 µl) (ni sledi hG-CSF)

Nanos 2: BL21(DE3) pET3a-hG-CSF induciran z IPTG pri 25^0C (10 µl) (rahla sled hG-CSF)

Nanos 3: BL21(DE3) pET3a-hG-CSF neinduciran pri 42^0C (10 µl) (ni sledi hG-CSF)

Nanos 4: BL21(DE3) pET3a-hG-CSF induciran z IPTG pri 42^0C (10 µl) (pod 1 % hG-CSF)

Nanos 5: standard filgrastim 0.3 µg za Coomassie brilliant blue

Nanos 6: BL21(DE3) pET3a-Fopt5 neinduciran pri 25^0C (5 µl) (6 % hG-CSF)

Nanos 7: BL21(DE3) pET3a-Fopt5 induciran z IPTG pri 25^0C (5 µl) (nad 50% hG-CSF)

b) detekcija s protitelesi (Western blot); primarna zajčja protitelesa; sekundarna kozja protitelesa proti zajčjim IgG konjugirana s peroksidazo iz hrena, substrat β-naftol

Vzorci za detekcijo s protitelesi so nanešeni v enakih količinah in enakem zaporedju kot pri SDS-PAGE (Slika 3a) z izjemo standarda, katerega nanos je bil 0.08 µg.

Slika 4: SDS-PAGE (4 % zgoščevalni, 15 % ločitveni; barvano s Coomassie brilliant blue) vzorcev proteinov inducirane kulture produkcijskega seva

E. coli BL21 (DE3) z ekspresijskim plazmidom pET3a pri 25° C. Kulture so bile gojene v GYSP/amp100 in LYSP/amp100 mediju.

Legenda:

Nanos 1: LMW (BioRad)

Nanos 2: BL21 (DE3) pET3a/P-Fopt5, kultura gojena v LYSP/amp100; (60% hG-CSF)

Nanos 3: BL21 (DE3) pET3a/P-Fopt5, kultura gojena v LYSP/amp100; (nad 54% hG-CSF)

Nanos 4: rhG-CSF (0.6 μg)

Nanos 5: rhG-CSF (1.5 μg)

Nanos 6: BL21 (DE3) pET3a/P-Fopt5, kultura gojena v GYSP/amp100 (4 μl); (55% hG-CSF)

Nanos 7: BL21 (DE3) pET3a/P-Fopt5, kultura gojena v GYSP/amp100 (5μl); (52% hG-CSF)

Primeri:

Primer 1: priprava optimalnega gena: Fopt5

Primer 1a: Predpriprava gena in plazmidov

Gen za hG-CSF gen smo namnožili iz BBG13 (R&D) z metodo PCR, s katero smo z začetnimi oligonukleotidi vnesli na začetek in konec gena tudi restrikcijski mesti Ndel in BamHI. Gen smo nato vključili v plazmid pCytexΔH,H (glej opis v nadaljevanju) med restrikcijski mesti Ndel in BamHI, v katerem smo tudi izvedli vse stopnje optimizacije gena za ekspresijo v *E. coli*.

V predpripravi smo iz gena za hG-CSF odstranili EcoRV mesto (oligo M20z108) zato, da smo si zagotovili možnost uvajanja točkovnih (posameznih) mutacij z oligonukleotidno usmerjeno mutagenezo v plazmidu pCytexΔH,H s kitom Transformer™ Site-Directed Mutagenesis Kit (Clontech). V plazmidu pCytexΔH,H-G-CSF je bilo tako mogoče uporabljati selekcijo mutant preko restrikcijskih mest EcoRI/EcoRV.

Izhodni plazmid pCYTEXP1 (Medac, Hamburg) smo preoblikovali tako, da je bila ekspresija konstitutivna, kar pomeni, da smo izrezali del gena za cI857 represor

med obema restrikcijskima mestoma HindIII. Plazmid, ki smo ga tako dobili, smo imenovali pCytexΔH,H.

Oligonukleotid za odstranitev EcoRV mesta iz nativnega gena za hG-CSF:

M20z108 5' -CCT GGA AGG AAT ATC CCC CG-3'

Primer 1b: Optimizacija kodonov (Slika 1)

V prvi stopnji optimizacije smo pripravili sintetski del gena med restrikcijskima mestoma NdeI in SacI z lepljenjem petih kaset (A, B, C, D,E), ki so bile sestavljene iz komplementarnih oligonukleotidov. Ta sintetski del gena predstavlja segment I. S segmentom I smo nato nadomestili del nativnega hG-CSF gena med restrikcijskima mestoma NdeI in SacI. To smo naredili tako, da smo izrezali prvi del gena med restrikcijskima mestoma Nde I in SacI in ga nadomestili s sintetsko pripravljeno kaseto. Postopek je potekal v dveh korakih. Najprej smo ligirali kaseto A, ki se je prilepila na NdeI mesto in kaseto E, ki se je prilepila na SacI mesto. Po 16 urah na 16^0C smo ligacijsko zmes oborili z etanolom, da smo odstranili presežek (nevezanih) oligonukleotidov ter nato v drugem koraku dodali srednji del celotne kasete (kaseta B, C in D) iz treh parov predhodno zlepljenih komplementarnih oligonukleotidov ter ponovno ligirali 16 ur na 16^0C.

V drugi stopnji optimizacije smo z usmerjeno oligonukleotidno mutagenezo (TransformerTM Site-Directed Mutagenesis Kit (Clontech)) zamenjali dva, za *E. coli* najbolj kritična kodona, Arg148 in Gly150, ki se nahajata v segmentu III.

V tretji stopnji optimizacije smo na podoben način kot v prvi, vendar brez vmesne etanolne precipitacije, pripravili segment IV, ki predstavlja zadnji del gena med restrikcijskima mestoma NheI in BamHI iz dveh parov komplementarnih oligonukleotidov (kaseta F in G).

V četrti stopnji optimizacije smo z usmerjeno oligonukleotidno mutagenezo (TransformerTM Site-Directed Mutagenesis Kit (Clontech)) zamenjali neugoden kodon za Ile96 (ATA→ATT) (segment II) ter vnesli restrikcijsko mesto Apa I (Gly101 GGT→GGG), ki se nahaja na 3' koncu segmenta II.

Apa I mesto smo nato uporabili v peti stopnji optimizacije zato, da smo nativni gen med Sac I in Apa I zamenjali s sintetsko DNA (segment II). Ta sintetska DNA je

sestavljena iz treh parov komplementarnih oligonukleotidov (kaseta H, I in J), ki smo jo izvedli na podoben način kot v prvi stopnji s kasnejšim dodatkom kasete I.

1.stopnja optimizacije:

komplementarni pari oligonukleotidov (Nde I – Sac I; segment I na Sliki 1):

Kaseta A: sestavljena iz komplementarnih oligonukleotidov zg1os1 in sp1os2:

zg1os1 5' TAT GAC ACC ACT GGG TCC AGC TTC TTC TCT GCC GCA AAG 3'

sp1os2 5' GCA GAG AAG AAG CTG GAC CCA GTG GTG TCA 3'

Kaseta B: sestavljena iz komplementarnih oligonukleotidov zg2os3 in sp2os4:

zg2os3 5' CTT TCT GTT GAA ATG TTT AGA ACA AGTTCG TAA AAT TCA AG 3'

sp2os4 5' GAA CTT GTT CTA AAC ATT TCA ACA GAA AGC TTT GCG 3'

Kaseta C: sestavljena iz komplementarnih oligonukleotidov zg3os5 in sp3os6:

zg3os5 5' GTG ATG GTG CAG CTT TAC AAG AAA AAC TGT GTG 3'

sp3os6 5' GTT TTT CTT GTA AAG CTG CAC CAT CAC CTT GAA TTT TAC 3'

Kaseta D: sestavljena iz komplementarnih oligonukleotidov zg4os7 in sp4os8:

zg4os7 5' CAA CTT ATA AAC TGT GTC ATC CAG AAG AAC TGG TTC TGT TAG 3'

sp4os8 5' CAG TTC TTC TGG ATG ACA CAG TTT ATA AGT TGC ACA CA 3'

Kaseta E: sestavljena iz komplementarnih oligonukleotidov zg5os9 in sp5os10:

zg5os9 5' GTC ATT CTC TGG GTA TTC CGT GGG CTC CTC TGA GCT 3'

sp5os10 5' CAG AGG AGC CCA CGG AAT ACC CAG AGA ATG ACC TAA CAG AAC 3'

2. stopnja optimizacije: oligonukleotidi za zamenjavo najbolj kritičnih kodonov z usmerjeno oligonukleotidno mutagenezo

zamenjava Arg 148 (CGG – CGT) in Gly 150 (GGA – GGT)

m38os16

5' CTC TGC TTT CCA GCG CCG TGC AGG TGG GGT CCT GGT TG 3'

3.stopnja optimizacije: komplementarni pari oligonukleotidov (Nhe I – BamH I; segment IV na Sliki 1):

Kaseta F: sestavljena iz komplementarnih oligonukleotidov zg6os11 in sp6os12:

zg6os11 5' CTA GCC ATC TGC AAT CTT TTC TGG AAG TTA G 3'

sp6os12 5' ACG ATA GCT AAC TTC CAG AAA AGA TTG CAG ATG G 3'

Kaseta G: sestavljena iz komplementarnih oligonukleotidov zg7os13 in sp7os14:

zg7os13 5' CTA TCG TGT TCT GCG TCA TCT GGC TCA GCC GTG ATA AG 3'

sp7os14 5' GAT CCT TAT CAC GGC TGA GCC AGA TGA CGC AGA AC 3'

4. stopnja optimizacije: oligonukleotidi za vpeljavo Apa I (Gly101 GGT →GGG), in zamenjavo neugodnega kodona Ile96 z usmerjeno oligonukleotidno mutagenezo vpeljava Apa I (Gly101 GGT →GGG), in zamenjava Ile 96 (ATA – ATT):

Apalos15

5' GCC CTG GAG GGG ATT TCC CCC GAG TTG GGG CCC ACC TTG GAC AC 3'

5.stopnja optimizacije: komplementarni pari oligonukleotidov (Sac I – Apa I;segment II na Sliki 1):

Kaseta H: sestavljena iz komplementarnih oligonukleotidov zg8os18 in sp8os19:

zg8os18 5' CCT GTC CGA GCC AGG CGC TGC AGC TGG CAG GCT GCC TGA G 3'

sp8os19 5' CCT GCC AGC TGC AGC GCC TGG CTC GGA CAG GAG CT 3'

Kaseta I: sestavljena iz komplementarnih oligonukleotidov zg9os20 in sp9os21:

zg9os20 5' CCA ACT GCA TAG CGG TCT GTT TCT GTA TCA GGG TCT GCT G 3'

sp9os21 5' CTG ATA CAG AAA CAG ACC GCT ATG CAG TTG GCT CAG GCA G 3'

Kaseta J: sestavljena iz komplementarnih oligonukleotidov zg10os22 in sp10os23:
zg10os22 5' CAG GCG CTG GAA GGC ATT TCC CCG GAA CTG GGG CC 3'
sp10os23 5' CCA GTT CCG GGG AAA TGC CTT CCA GCG CCT GCA GCA GAC C 3'

Primer 2: Ekspresija sintetskega gena za hG-CSF v *E. coli*

Optimiran gen Fopt5 smo iz plazmida pCyΔH,H izrezali s restrikcijskima encimoma NdeI in BamHI ter ga subklonirali v končni ekspresijski plazmid pET3a (Novagen, Madison USA) ter transformirali v produkcijski sev *E. coli* BL21 (DE3).

Kulture smo pripravili na stresalniku 24 ur na 160 rpm pri 25^{0}C oziroma 15 ur 42^{0}C:

- v LBG10/amp100 mediju (10 g/l tripton, 5g/l kvasni ekstrakt, 10 g/l NaCl, 10 g/l glukoze, 100 mg/L ampicilina). Indukcijo smo izvedli z dodatkom IPTG v medij do končne koncentracije 0.4 mM.

Kulture smo pripravili na stresalniku 24 ur na 160 rpm pri 25^{0}C:

- v GYSP/amp100 mediju (20 g/l fiton, 5g/l kvasni ekstrakt, 10 g/l NaCl, 10 g/l glukoze, kovine v sledovih, 100 mg/L ampicilina). Indukcijo smo izvedli z dodatkom IPTG v medij do končne koncentracije 0.4 mM.
- v LYSP/amp100 mediju (20 g/l fiton, 5g/l kvasni ekstrakt, 10 g/l NaCl, 6 g/l glicerola, 4 g/l laktoze, kovine v sledovih, 100 mg/L ampicilina). Indukcijo smo v tem primeru izvedli z dodatkom laktoze v medij.

Inokulum smo pripravili v LBG/amp100 mediju (10 g/L tripton, 5 g/L kvasni ekstrakt, 10 g/L NaCl, 2.5 g/L glukoze) in 100 mg/L ampicilina pri 25^{0}C, 160 rpm prekonočno.

Za analizo smo centrifugirali po 8 ml kulture na 5000 rpm. Pelete smo nato resuspendirali v 10 mM TrisHCl/pH=8.0 tako, da smo dodali 0.66 ml pufra preračunano na 1 enoto OD_{600nm}. Tako smo izenačili količino nanosov celokupnih proteinov, ker končni OD_{600nm} kultur v navedenih primerih niso bili enaki. Vzorce smo zmešali v razmerju 3:1 s 4x SDS - vzorčnim pufrom z DTT (pH=8.7) in segrevali 10 minut na 95^{0}C, odcentrifugirali ter tako pripravljene nanesli na gel.

Deleži akumuliranega hG-CSF, ki se izloča v obliki inkluzijskih teles, so za nativni in optimirani gen opisani v Tabeli 1.

Tabela 1: Primerjava nivojev akumulacije hG-CSF pri nativnem in optimiranem genu (Fopt5)

		delež (%) hG-CSF od celotnih proteinov		
Ekspresijski sistem	pogoji gojenja in indukcije	nativen gen za hG-CSF		optimiran gen Fopt5
	temperatura gojenja	25° C	42° C	25° C
pET3a / *E. coli* BL21 (DE3)	gojišče LBG10/amp100 0.4 mM IPTG	sled	< 1 %	> 40 %
pET3a / *E. coli* BL21 (DE3)	gojišče GYSP/amp100	< 1 %	< 1 %	> 52 %
pET3a / *E. coli* BL21 (DE3)	gojišče LYSP/amp100	< 1 %	< 1 %	> 52 %

Navedene vrednosti za vsebnost hG-CSF so pri Fopt5 dobljene z denzitometrično analizo SDS-PAGE gelov obarvanih s Coomassie brilliant blue (Slika 3a in Slika 4) oziroma v primeru neoptimiranega gena po detekciji s protitelesi (Slika 3b). Relativni delež je bil pri oceni ekspresije Fopt5 določen s profilno analizo (program Molecular analyst; BioRad) gelov na aparatu Imaging densitometer Model GS670 (BioRad).

Opis DNA zaporedij

<110> Lek farmacevtska družba d. d.

<120> Sintetski gen za humani granulocitne kolonije stimulirajoči dejavnik za ekspresijo v *E. coli*

<160> 2

<210> SEQ ID NO: 1

<211> 525 baznih parov

<212> DNA

<213> sintetsko zaporedje

<220> gen

<400> SEQ ID NO: 1

```
atgacaccac tgggtccagc ttcttctctg ccgcaaagct ttctgttgaa atgtttagaa    60
caagttcgta aaattcaagg tgatggtgca gctttacaag aaaaactgtg tgcaacttat   120
aaactgtgtc atccagaaga actggttctg ttaggtcatt ctctgggtat tccgtgggct   180
cctctgagct cctgtccgag ccaggcgctg cagctggcag gctgcctgag ccaactgcat   240
agcggtctgt ttctgtatca gggtctgctg caggcgctgg aaggcatttc cccggaactg   300
gggcccacct tggacacact gcagctggac gtcgccgact ttgccaccac catctggcag   360
cagatggaag aactgggaat ggcccctgcc ctgcagccca cccagggtgc catgccggcc   420
ttcgcctctg ctttccagcg ccgtgcaggt ggggtcctgg ttgctagcca tctgcaatct   480
tttctggaag ttagctatcg tgttctgcgt catctggctc agccg                   525
```

<210> SEQ ID NO: 2

<211> 528 baznih parov

<212> DNA

<213> sintetsko zaporedje

<220> gen

<400> SEQ ID NO: 2

```
atgacaccac tgggtccagc ttcttctctg ccgcaaagct ttctgttgaa atgtttagaa    60
caagttcgta aaattcaagg tgatggtgca gctttacaag aaaaactgtg tgcaacttat   120
aaactgtgtc atccagaaga actggttctg ttaggtcatt ctctgggtat tccgtgggct   180
cctctgagct cctgtccgag ccaggcgctg cagctggcag gctgcctgag ccaactgcat   240
agcggtctgt ttctgtatca gggtctgctg caggcgctgg aaggcatttc cccggaactg   300
gggcccacct tggacacact gcagctggac gtcgccgact ttgccaccac catctggcag   360
cagatggaag aactgggaat ggcccctgcc ctgcagccca cccagggtgc catgccggcc   420
ttcgcctctg ctttccagcg ccgtgcaggt ggggtcctgg ttgctagcca tctgcaatct   480
tttctggaag ttagctatcg tgttctgcgt catctggctc agccgtga                528
```

Lek farmacevtska družba d. d.

Patentni zahtevki

1. DNA zaporedje, ki kodira za hG-CSF, označeno s tem, da vsebuje zaporedje nukleotidov po SEQ ID NO: 1.
2. DNA zaporedje, označeno s tem, da vsebuje zaporedje nukleotidov iz skupine, ki zajema delno zaporedje SEQ ID NO: 1 in nukleinske kisline, ki hibridizirajo z zaporedjem po SEQ ID NO: 1 pri zaostrenih pogojih.
3. Ekspresijski plazmid, označen s tem, da vsebuje DNA zaporedje po zahtevku 1 in plazmidni vektor.
4. Ekspresijski plazmid, označen s tem, da vsebuje DNA zaporedje po zahtevku 2 in plazmidni vektor.
5. Ekspresijski plazmid po zahtevkih 3 in 4, označen s tem, da je plazmidni vektor izbran iz skupine pET vektorjev.
6. Ekspresijski sistem za ekspresijo DNA zaporedja po zahtevku 1, označen s tem, da vsebuje DNA zaporedje, plazmidni vektor in produkcijski sev *E. coli.*
7. Ekspresijski sistem za ekspresijo DNA zaporedja po zahtevku 2, označen s tem, da vsebuje DNA zaporedje, plazmidni vektor in produkcijski sev *E. coli.*
8. Ekspresijski sistem po zahtevkih 6 in 7, označen s tem, da je plazmidni vektor izbran iz skupine pET vektorjev.
9. Ekspresijski sistem po zahtevkih 6 in 7, označen s tem, da je produkcijski sev *E. coli* BL21 (DE3).
10. Priprava DNA zaporedja po zahtevku 1, označena s tem, da vključuje postopke, ki so izbrani iz skupine, ki zajema:

    -zamenjavo nekaterih za *E. coli* neugodnih kodonov s kodoni, ki so za *E. coli* ugodni,

    -zamenjavo nekaterih GC bogatih regij z AT bogatimi regijami.

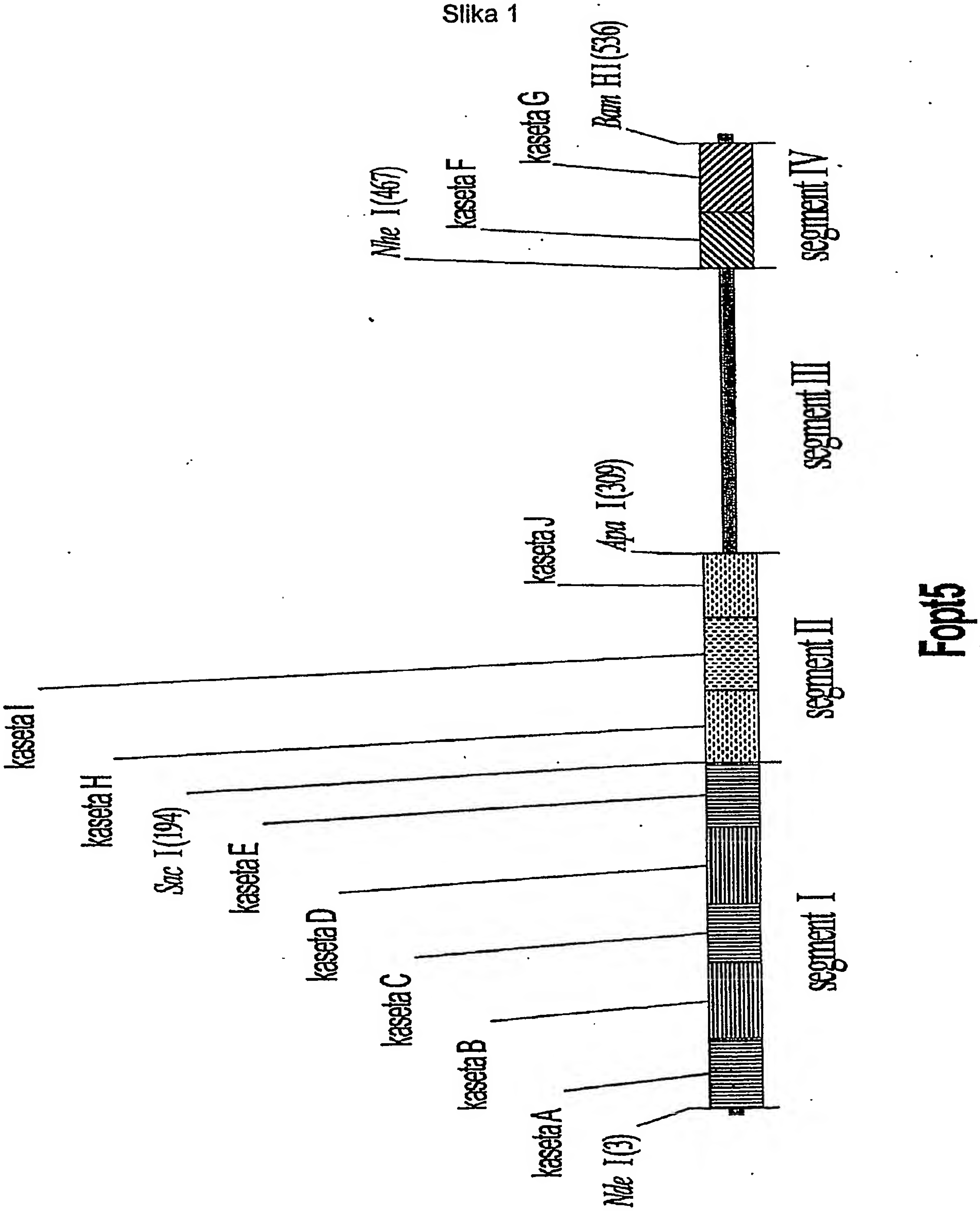
    in nadalje zajema popolnoma nespremenjen del nativnega zaporedja za hG-CSF.
11. Priprava DNA zaporedja po zahtevku 10, ki nadalje ne vključuje sprememb v področjih iz skupine, ki zajema: področja iniciacije translacije, področja vezavnega mesta za ribosom ter področja v razdalji med start kodonom in vezavnim mestom za ribosom.
12. Ekspresija DNA zaporedja po zahtevku 1 v *E. coli.*
13. Ekspresija DNA zaporedja po zahtevku 2 v *E. coli.*

Lek farmacevtska družba d. d.

Izvleček

Izum se nanaša na sintetski gen za hG-CSF, ki omogoča ekspresijo v *E. coli* z nivojem ekspresije več kot 52% rekombinantnega hG-CSF glede na celotne proteine po ekspresiji.

Slika 1



Lek farmacevtska družba d. d.

21-07-2002

Slika 2

a)

ATGACCCCCCTGGGCCCTGCCAGCTCCCTGCCCCAGAGCTTCCTGCTCAAGTG
CTTAGAGCAAGTGAGGAAGATCCAGGGCGATGGCGCAGCGCTCCAGGAGAAGC
TGTGTGCCACCTACAAGCTGTGCCACCCCGAGGAGCTGGTGCTGCTCGGACAC
TCTCTGGGCATCCCCTGGGCTCCCCTGAGCTCCTGCCCCAGCCAGGCCCTGCA
GCTGGCAGGCTGCTTGAGCCAACTCCATAGCGGCCTTTTCCTCTACCAGGGGC
TCCTGCAGGCCCTGGAAGGGATATCCCCCGAGTTGGGTCCCACCTTGGACACA
CTGCAGCTGGACGTCGCCGACTTTGCCACCACCATCTGGCAGCAGATGGAAGA
ACTGGGAATGGCCCCTGCCCTGCAGCCCACCCAGGGTGCCATGCCGGCCTTCG
CCTCTGCTTTCCAGCGCCGGGCAGGAGGGGTCCTGGTTGCTAGCCATCTGCAG
AGCTTCCTGGAGGTGTCGTACCGCGTTCTACGCCACCTTGCGCAGCCC

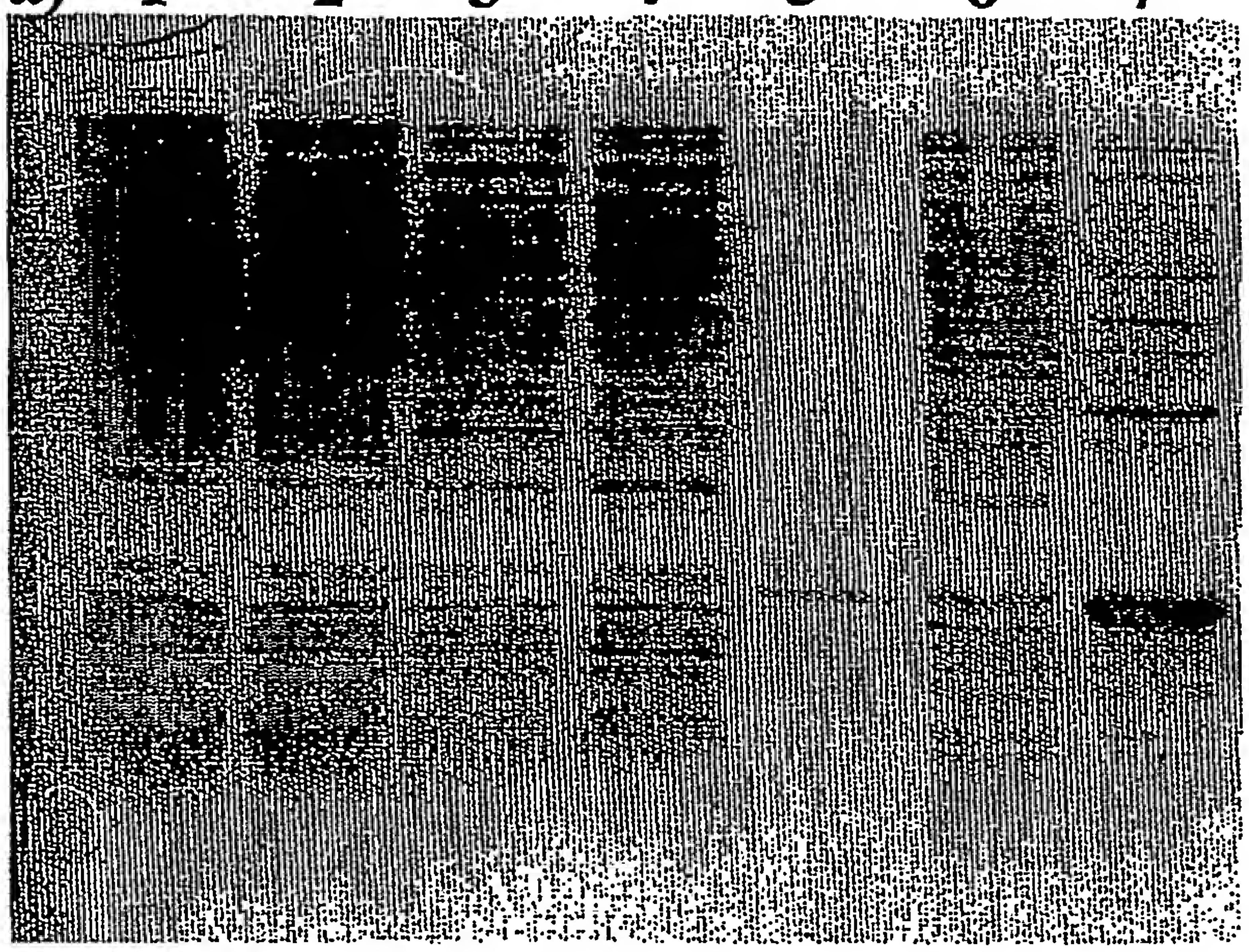
b)

ATGACACCACTGGGTCCAGCTTCTTCTCTGCCGCAAAGCTTTCTGTTGAAATG
TTTAGAACAAGTTCGTAAAATTCAAGGTGATGGTGCAGCTTTACAAGAAAAAC
TGTGTGCAACTTATAAACTGTGTCATCCAGAAGAACTGGTTCTGTTAGGTCAT
TCTCTGGGTATTCCGTGGGCTCCTCTGAGCTCCTGTCCGAGCCAGGCGCTGCA
GCTGGCAGGCTGCCTGAGCCAACTGCATAGCGGTCTGTTTCTGTATCAGGGTC
TGCTGCAGGCGCTGGAAGGCATTTCCCCGGAACTGGGGCCCACCTTGGACACA
CTGCAGCTGGACGTCGCCGACTTTGCCACCACCATCTGGCAGCAGATGGAAGA
ACTGGGAATGGCCCCTGCCCTGCAGCCCACCCAGGGTGCCATGCCGGCCTTCG
CCTCTGCTTTCCAGCGCCGTGCAGGTGGGGTCCTGGTTGCTAGCCATCTGCAA
TCTTTTCTGGAAGTTAGCTATCGTGTTCTGCGTCATCTGGCTCAGCCG

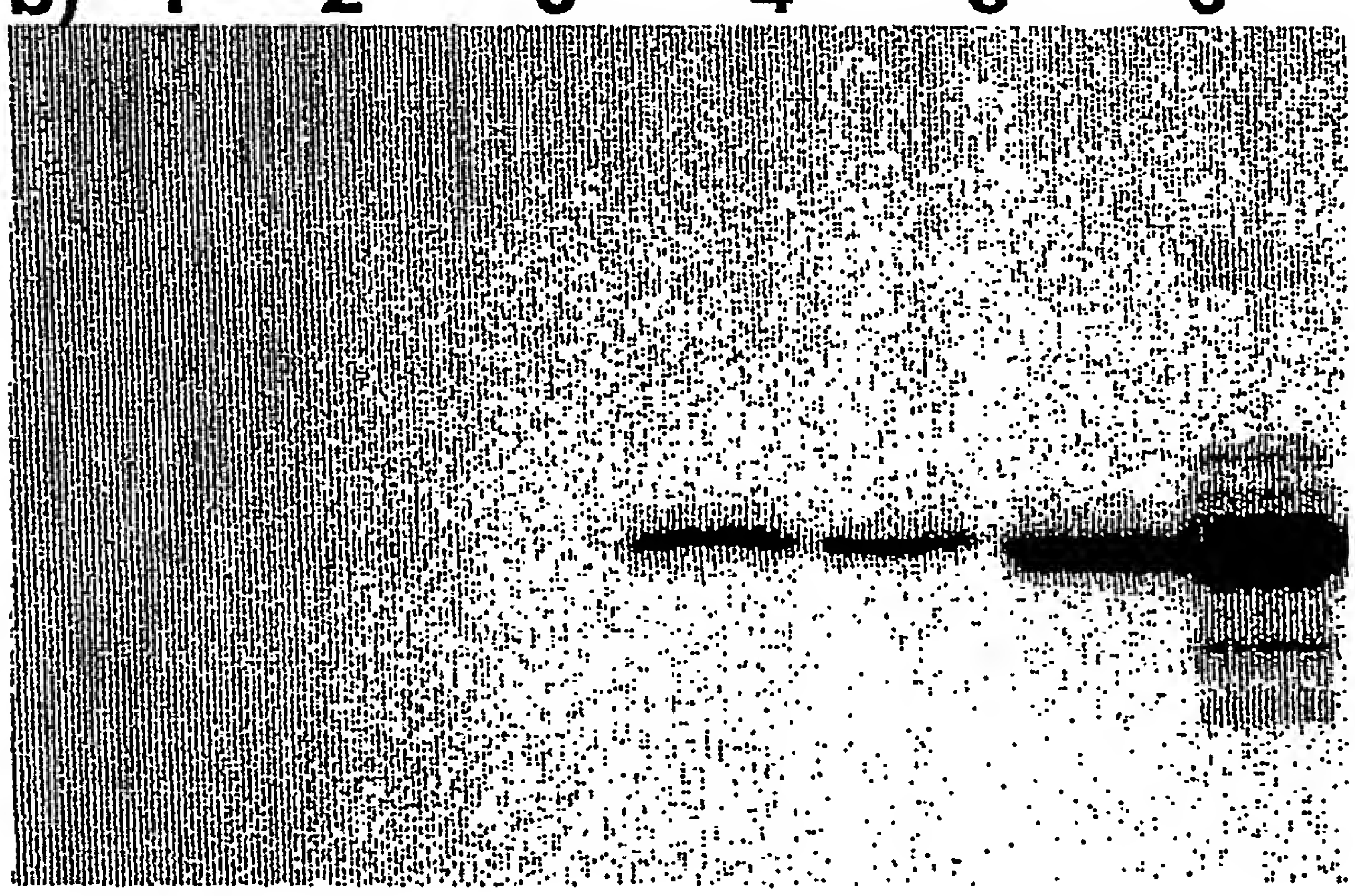
Lek farmacevtska družba d. d.

Slika 3

a) 1 2 3 4 5 6 7

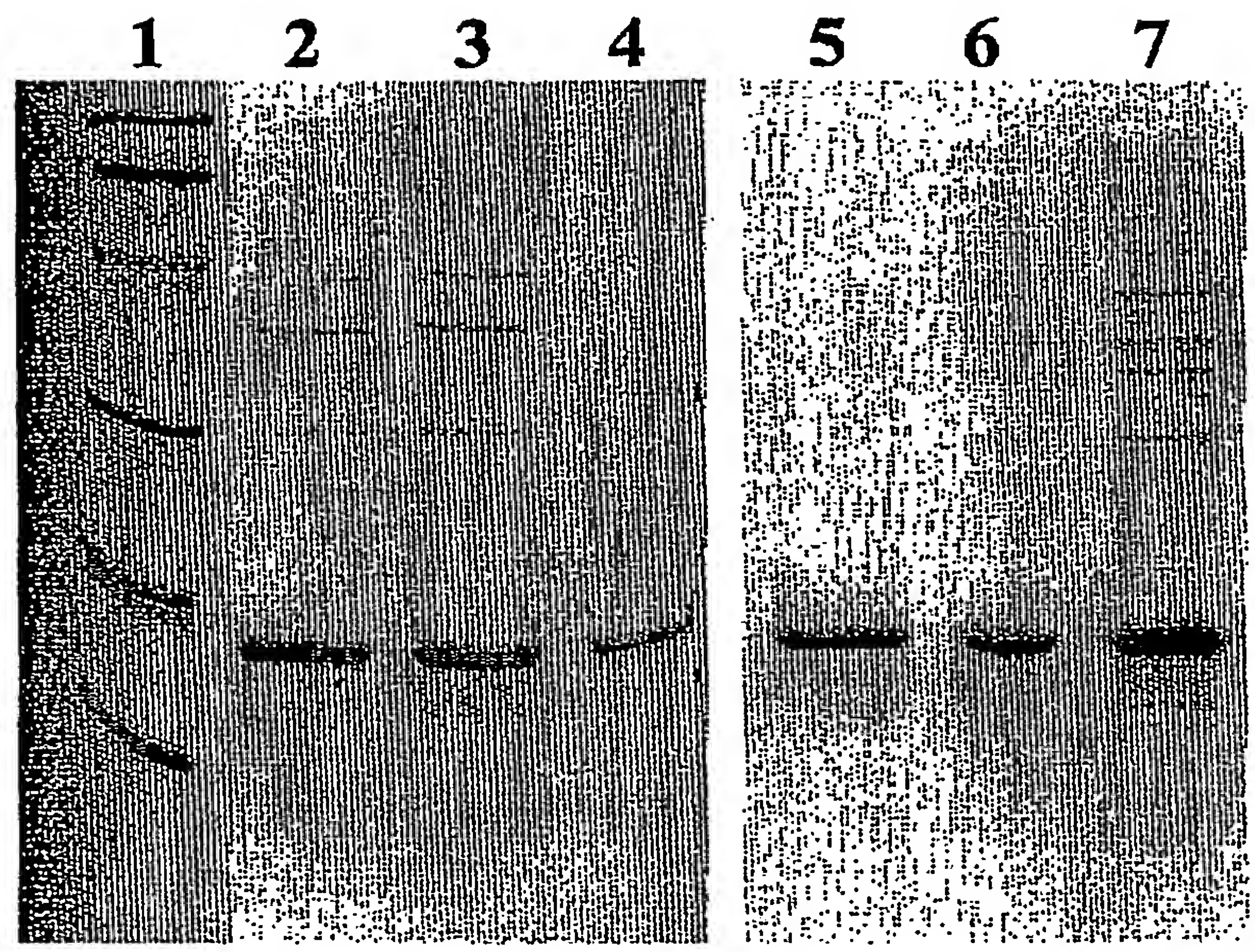


b) 1 2 3 4 5 6



Lek farmacevtska družba d. d.

Slika 4



Lek farmacevtska družba d. d.

REPUBLIC OF SLOVENIA

Ministry of Economic Affairs

SLOVENIAN INTELLECTUAL PROPERTY OFFICE

Certificate

Slovenian Intellectual Property Office hereby certifies that the document annexed hereto is a true copy of the patent application, as follows:

(22) Application Date:

31 July 2002

(21) Application No.:

P-200200188

(54) Title:

Synthetic gene coding for human granulocyte-colony stimulating factor for the expression in E. coli

Ljubljana, 3 June 2003

Janez Kukec-Mezek
Government Counsellor

L.S.
Republic of Slovenia
Ministry of Economic Affairs
Slovenian Intellectual Property Office
Ljubljana

Kotnikova 6, 1001 Ljubljana, POB 206, telephone: 01/478 3100, fax: 01/478 3111

REPUBLIC OF SLOVENIA
MINISTRY OF SCIENCE AND TECHNOLOGY

SLOVENIAN INTELLECTUAL PROPERTY OFFICE
1000 LJUBLJANA, Kotnikova 6

REQUEST FOR A PATENT GRANT

1. Address for correspondence: Lek Pharmaceuticals d.d. Intellectual Property Verovškova 57 1526 Ljubljana Slovenia Telephone: 01 580 2005 Fax: 01 568 2123 Code: bk / 910	Acknowledgement of the application *(for official use only)* Date of application receipt: 31 July 2002 Application number: P - 200200188 Stamp of the office and signature:
2. Applicant (Family name followed by given name and address; for a legal entity, full official designation Lek Pharmaceuticals d.d. Verovškova 57 1526 Ljubljana Slovenia	
3. Representative:	Registration No.:
4. Inventor (Family name followed by given name and address): Jevševar Simona Vodole 26 2000 Maribor	
5. Title of invention: Synthetic gene coding for human granulocyte-colony stimulating factor for the expression in *E. coli*	
6. Claimed priority right:	
7. Additional requests: ☐ application for a shortened duration patent ☐ preliminary publication after the expiry of ___months ☐ application is divided from the application no.:___	
8. Statements: ☐ statement of common representative	

REPUBLIC OF SLOVENIA
MINISTRY OF SCIENCE AND TECHNOLOGY

SLOVENIAN INTELLECTUAL
PROPERTY OFFICE
1000 LJUBLJANA, Kotnikova 6

9. Enclosures:

- x Description of the invention, having 17 pages
- x Patent claim (claims), having 1 page; number of claims: 13
- x Schemes (if required for patent description); number of sheets: 4
- x Abstract
- ☐ Voucher for the settlement of fees
- ☐ Declaration of depositing the biological material if it is an invention which cannot be described
- ☐ Authorisation to the representative
- ☐ General authorisation to the representative is deposited in the office under no. ....
- ☐ Declaration of priority right
- ☐ Information of additional applicants
- x Information of additional inventors
- ☐ Presentation of nucleotide or amino acid sequence in the description
- x Application was previously faxed or mailed in electronic form
- x Statement of the applicant on presentation of the sequence

Alenka Košak

Applicant's (representative's) family name
followed by given name and signature,

Information of additional inventors:

Viktor Menart
Logatec, Dol 3
1370 Logatec

Title of the invention

Synthetic gene coding for human granulocyte-colony stimulating factor for the expression in *E. coli*

Field of the invention

The present invention relates to synthetic gene coding for human granulocyte-colony stimulating factor (hG-CSF) which enables expression in *E. coli* with an of expression level being equal as or higher than 52% of the recombinant hG-CSF to the total proteins after expression.

hG-CSF belongs to a family of stimulating factors which regulate the differentiation and proliferation of hematopoetic mammalian cells. They have a major role in the neutrophil formation and are therefore suitable for use in medicine in the field of hematology and oncology.

Two forms of hG-CSF are currently available for clinical use on the market: lenograstim which is glycosylated and is obtained by the expression in mammalian cell line and filgrastim which is non-glycosylated and is obtained by the expression in the bacterium *Escherichia coli* (*E. coli*).

Summary of the invention

The essential feature of the present invention is that the use of synthetic gene coding for hG-CSF enables to attain a level of expression (accumulation) in *E. coli* being equal to or higher than 52% of recombinant hG-CSF regarding the total proteins in *E. coli*. The expression plasmid containing a strong T7 promoter is used for the expression. The synthetic gene coding for hG-CSF is constructed by using a complex combination of two methods which enable the construction of optimized synthetic gene (coding for hG-CSF) for its expression in *E. coli*. The first method includes the replacement of some rare *E. coli* codons which are unfavorable for expression in *E. coli* with E. coli preference codons for which are more favorable for the expression in E. coli. The second method includes the replacement of some GC rich regions with AT rich regions. Some parts of the synthetic gene of the present invention are constructed by using one of the two methods, for some parts the combination of the two methods is used, whereas some parts of the gene are not

changed. At the construction procedure of the synthetic gene coding for hG-CSF, which is also the subject of the present invention, the non coding regions are not changed. This means that there are no modifications in either the translation initiation region (TIR) or in the ribosome binding site (RBS), or in the region between the start codon and RBS.

Background of the invention

The impact of several successive rare codons such as arginine codons (AGG/AGA; CGA), leucine codon (CTA), isoleucine codon (ATA) and proline codon (CCC), on the level of translation and consecutively on the decrease of the amount and quality of the expressed protein in *E. coli* are described in Kane JF, Current Opinion in Biotechnology, 6:494-500 (1995). There is a similar impact of individual rare codons if they occur in different parts of the gene.

The GC rich regions also have impact on the translational efficiency in *E. coli* if a stable double stranded RNA is formed in the mRNA secondary structure. This impact is the highest when the GC rich regions of mRNA are found either in the RBS, or in the direct proximity of the RBS or also in the direct proximity of the start codon (Makrides SC, Microbiological Reviews, 60:512-538 (1996); Baneyx F, Current Opinion in Biotechnology, 10:411-421 (1999)).

There are known several methods for the prediction of the secondary structure and calculating minimal free energy of individual RNA molecule which is supposed to be the basic rule for the most stable / most probable structure (SantaLucia J Jr and Turner DH, Biopolymers, 44:309-319 (1997)). The reliable algorithms for the prediction of the correct secondary structure are not known with the exception of some cases. There has been no evidence for the quantitative correlation with the expression level (Smit MH and van Duin JJ. Mol. Biol., 244, 144-150 (1994)). It is still impossible to predict the tertiary structures of RNA (Tinoco I and Bustamante C, J. Mol. Biol, 293:271-281 (1999)).

The increase of the expression level after the optimization of DNA sequence in the TIR region, in the RBS region and in the region between the start codon and the RBS region is described in McCarthy JEG and Brimacombe R, Trends Genet 10:402-

407 (1994). In this case the expression level increased due to more efficient translation initiation and its smooth continuation in the mRNA coding region.

The production of adequate amounts of hG-CSF for performing the *in vitro* biological studies by expression in *E. coli* is described in Souza LM et al, Science 232:61-65 (1986) and in Zsebo KM et al, Immunobiology 172:175-184 (1986). The hG-CSF expression level was lower than 1%.

The patent US4810643 discloses the use of synthetic gene coding for hG-CSF which was first of all constructed on the basis of replacement of *E. coli* rare codons with the *E. coli* preference codons. The combination with thermoinducible phage lambda promoter led to the expression level of 3 to 5% of hG-CSF regarding the total cellular proteins. This level was not sufficient for the economical large-scale production of hG-CSF.

8-10% accumulation of hG-CSF to total cellular proteins was reached by changing the first four codons in the 5' end region of hG-CSF as is described in Wingfield P et al, Biochem. J, 256:213-218 (1988).

The expression of hG-CSF in *E. coli* with the yield up to 17% of hG-CSF to total cellular bacterial proteins is described in Devlin PE et al, Gene 65:13-22 (1988). Such yield was reached with partial optimization of DNA sequence in the 5' end of the G-CSF coding region (codons coding for the first four amino acids) whereby the GC region was replaced with AT region and a relatively strong lambda phage promoter was used. This expression level is not very high what leads to lower production yields and is less economical in the large-scale production.

The use of synthetic gene and the expression level of about 30% are described in Kang SH et al, Biotechnology letters, 17(7):687-692 (1995). This level was attained by the introduction of *E. coli* preference codons, by the modifications in the TIR region and with the additional modifications of codon sets whereby the 3' end of the gene was not essentially changed. Thus, for attaining the stated expression level the changes of the gene in the TIR region were needed and the expression level did not exceed 30%. The patent US5840543 describes the synthetic gene coding for hG-CSF which was constructed by the introduction of AT rich regions at the 5' end of the gene and with the replacement of *E. coli* rare codons with *E. coli* preference codons. Under the control of the Trp promoter the expression with the

yield of 11% hG-CSF to total cellular proteins was reached. On the other hand, the addition of leucine and threonine or their combination into the fermentation medium (where the bacteria were cultivated) led to the accumulation of up to 35% of hG-CSF regarding total cellular proteins. Such expression level was therefore reached by the addition of amino acids into the fermentation medium what is an additional cost in the process for production of hG-CSF and is not economical for the industrial production. Only optimization of the gene coding for hG-CSF did not enable a higher expression level of hG-CSF.

The highest accumulation of hG-CSF regarding total cellular proteins found in the prior art is described in v Jeong et al, Protein Expression and Purification 23,:311-318 (2001) and is 48%. Such accumulation was obtained by the changes in the N-terminal end and by the induction with 1 mM IPTG.

In general, there are no reports on possible predictions of the expression level of native human genes in prokaryotic organisms, e.g. bacterium *E. coli.* The described expression levels are relatively low or difficult to detect even when the expression plasmids with strong promoters, e.g. from lambda or T7 phage are used. From the prior art literature it can be gathered that many parameters (rare codons or their clustering; GC base pairs rich regions, unfavorable mRNA secondary structures, unstable mRNA) have an impact on the accumulation of a human protein in *E. coli.*

Until now there has been no entirely developed rule known on how to combine the codons in order to obtain the secondary or tertiary mRNA structures which are optimal for expression. Although there exist some mathematical and structural models for predicting and thermodynamical stability of secondary structures, but they are too unreliable to predict the secondary structures. On the other hand, there are no such models for predicting the tertiary structures. These currently accessible models therefore do not enable the prediction of the impact of the codons on the expression level.

There are no reports in either the patent or the scientific literature on the more efficient way for solving the problem of low expression level of the native gene coding for hG-CSF in *E. coli.*

Description of the invention

It has been found that the problem with the low expression level of the native gene coding for hG-CSF in *E. coli* can be solved by the optimization of the native gene coding for hG-CSF leading to the construction of the synthetic gene coding for hG-CSF. In comparison with the data described in the art, surprisingly high expression level can be obtained.

The term 'hG-CSF', as used herein, refers to human granulocyte-colony stimulating factor, comprising the recombinant hG-CSF obtained by the expression in *E. coli.*

The synthetic gene encoding hG-CSF of the present invention was obtained by introducing changes in the nucleotide sequence of the gene encoding the native hG-CSF. Thus the amino acid sequence was not changed and remained identical to the native hG-CSF.

The present invention further comprises the expression of the synthetic gene in *E. coli* and the level of expression of the synthetic gene.

The term 'expression level', as used herein, refers to the proportion of hG-CSF obtained after the heterologous expression of the gene encoding hG-CSF regarding the total cellular proteins after expression.

The term 'heterologous expression', as used herein, refers to the expression of the genes which are foreign to the organism in which the expression occurs.

The term 'homologous expression', as used herein, refers to the expression of the genes which are proper to the organism in which the expression occurs.

The term 'preference codons', as used herein, refers to the codons used by an individual organism (e.g. *E. coli*) for the production of most mRNA molecules. The organism uses these codons for expressing genes with high homologous expression.

The term 'rare codons' as used herein, refers to the codons used by an individual organism (e.g. *E. coli*) only for expressing genes with low expression level. These codons are rarely used in the organism (low homologous expression).

The term 'GC rich regions', as used herein, refers to the regions in the gene where the bases guanine (G) and cytosine (C) prevail.

The term 'AT rich regions', as used herein, refers to the regions in the gene, where the bases adenine and thymine prevail.

The term 'synthetic gene', as used herein, refers to the gene which differs from the native gene only in the nucleotide sequence whereby the amino acid sequence remains unchanged. The synthetic gene is obtained by the techniques of the recombinant DNA technology.

The term 'native gene', as used herein, refers to a gene which is not modified by using the techniques of the recombinant DNA technology.

The term 'segment', as used herein, refers to the parts of the genes which are bounded by single restriction sites on both ends. These sites serve as subcloning sites for the synthetically constructed parts of the gene.

The term 'segment I', as used herein, refers to the 5' end of the gene encoding hG-CSF between the restriction sites Nde I (3) and Sac I (194), i.e. 191 bp long sequence which was de novo synthesized.

The term 'segment II', as used herein, refers to the part of the gene for hG-CSF between restriction sites Sac I (194) and Apa I (309), i.e. 115 bp long central part of the gene which was de novo synthesized.

The term 'segment III', as used herein, refers to the part of the gene for hG-CSF between restriction sites Apa I (309) in Nhe I (467), i.e. 158 bp long part of the gene where the native DNA sequence for hG-CSF is preserved with the exception of Arg148 and Gly150.

The term 'segment IV', as used herein, refers to the 3' terminal end of the gene encoding hG-CSF between restriction sites Nhe I (467) and BamH I (536), i.e. 69 bp long terminal part of the gene which was de novo synthesized.

The synthetic gene encoding hG-CSF of the present invention is constructed by the combination of the following methods:

- replacement of the E. coli rare codons with E. coli preference codons: in the segment II (between restriction sites Sac I (194) and Apa I (309)) and in the segment IV (between restriction sites Nhe I (467) and BamH I(536))
- replacement of some GC rich regions with AT rich regions, whereat the rarest E. coli codons are replaced, but mostly not with the E. coli preference codons: in the segment I (between restriction sites Nde I (3) and Sac I (194)).

- completely unchanged native sequence of 46 codons (between Pro102 and Arg147) in the segment III.
- elimination of two E. coli rare codons (Arg148 and Gly150) at the terminal end of the segment III.

Optimization of the gene coding for hG-CSF of the present invention does not include changes in the TIR, RBS and in the regions between the start codons and RBS.

The synthetic gene of the present invention encoding hG-CSF enables expression of the constructed synthetic gene encoding hG-CSF with the expression level in E. coli equal to or higher than 52%. Furthermore, the expression level of about 55% or even about 60% can also be obtained. High expression level of the synthetic gene coding for hG-CSF of the present invention enables high yields of hG-CSF production, faster and simpler purification and isolation of heterologous hG-CSF, easier in-process control, and the whole production process is more economical. Therefore, the efficient production of hG-CSF in industrial scale is enabled. The produced hG-CSF is suitable for clinical use in medicine.

The construction of the synthetic gene of the present invention begins with the initial preparation of the hG-CSF native gene and of the plasmids. Gene coding for native hG-CSF can be of human origin, but the same principle can be used for every gene which is homologous in the regions which comprise single restriction sites which are used for subcloning of de novo synthesized gene segments. The plasmid for mutagenesis was chosen according to its ability to be capable of enabling the successive introduction of point mutations. Selection of enrichment of the plasmids containing desired mutation was obtained by using an additional selection primer that changed unique restriction site EcoRI into EcoRV or vice-versa (Transformer™ Site-Directed Mutagenesis Kit (Clontech)). The gene and the plasmid are constructed in such a way that the introduction of point mutation by cassette mutagenesis is possible.

After the initial preparation of native gene coding for hG-CSF and of plasmids the optimization of the native gene coding for hG-CSF is performed. This means that the synthetic gene coding for hG-CSF is constructed. The optimization begins with the division of the native gene coding for hG-CSF into four (I, II, III in IV) segments,

which are or will be separated with single restriction sites after the oligonucleotide mutagenesis and in the individual segments the changes are introduced. In some individual segments the changes in the gene sequence are introduced whereas in certain segments the gene is not changed (Figure 1). The obtained optimized synthetic gene coding for hG-CSF therefore consists of partially preserved native sequence (segment III) and of 5' and 3' coding regions which are synthesized de novo (segments I, II in IV).

The changes in the individual segments:

Segment I: Replacement of E. coli rare codons with *E. coli* preference codons and replacement of GC rich regions with AT rich regions

Thr2 (ACC→AC**A**), Pro3 (CCC→CC**A**), Gly5 (GGC→GG**T**) Pro6 (CCT→CC**A**), Ala7 (GCC→GC**T**), Ser8 (AGC→**TCT**), Ser9 (TCC→TC**T**), Pro11 (CCC→CC**G**), Gln12 (CAG→CA**A**), Phe14 (TTC→TT**T**), Leu16 (CTC→**TTG**), Lys17 (AAG →AA**A**), Cys18 (TGC→TG**T**), Glu20 (GAG→GA**A**), Val22 (GTG →GT**T**), Arg23 (AGG→**CGT**), Lys24 (AAG→AA**A**) Ile25 (ATC→AT**T**), Gln26 (CAG→CA**A**), Gly27 (GGC→GG**T**), Gly29 (GGC→GG**T**), Ala31 (GCG→GC**T**), Leu32 (CTC→**T**T**A**), Gln33 (CAG→CA**A**), Glu34 (GAG→GA**A**), Lys35 (AAG →AA**A**), Ala38 (GCC→GC**A**), Thr39 (ACC→AC**T**), Tyr40 (TAC→TA**T**), Lys41 (AAG→AA**A**), Cys43 (TGC→TG**T**), His44 (CAC→CA**T**), Pro45 (CCC→CC**A**), Glu46 (GAG→GA**A**), Glu47 (GAG→GA**A**), Val49 (GTG→GT**T**), Leu51 (CTC→**T**T**A**), Gly52 (GGA→GG**T**), His53 (CAC→CA**T**), Gly56 (GGC→GG**T**), Ile57 (ATC→AT**T**), Pro58 (CCC→CC**G**), Pro61 (CCC→CC**T**)

Segment II: Replacement of *E. coli* rare codons with *E. coli* preference codons.

Cys65 (TGC→TG**T**), Pro66(CCC→CC**G**), Ala69 (GCC→GC**G**), Leu76 (TTG→**C**TG), Leu79 (CTC→CT**G**), Gly82 (GGC→GG**T**), Leu83 (CTT→CT**G**), Phe84 (TTC→TT**T**), Leu85 (CTC→CT**G**), Tyr86 (TAC→TA**T**), Gly88 (GGG→GG**T**), Leu89 (CTC→CT**G**), Ala92 (GCC→GC**G**), Gly95 (GGG→GG**C**), Ile96 (ATA→AT**T**), Pro98 (CCC→CC**G**), Glu99 (GAG→GA**A**), Leu100 (TTG→**C**TG), Gly101 (GGT→GG**G**)

Segment III: Replacement of two *E. coli* rare codons situated just before the restriction site NheI

Arg 148 (CGG →CG**T**), Gly150 (GGA→GG**T**)

Segment IV: Replacement of a long cluster of *E. coli* rare codons at the terminal end of the gene with *E. coli* preference codons.

Gln159 (CAG→CA**A**), Ser160 (AGC→**TCT**), Phe161 (TTC→TT**T**), Glu163 (GAG→GA**A**), Val164 (GTG→GT**T**), Ser165 (TCG→**AGC**), Tyr166 (TAC→TA**T**), Arg167 (CGC→CG**T**), Leu169 (CTA→CT**G**), Arg170 (CGC→CG**T**), His171 (CAC→CA**T**), Leu172 (CTT→CT**G**), Ala173 (GCG→GC**T**), Pro175 (CCC→CC**G**)

After the construction of the synthetic gene coding for hG-CSF the optimized synthetic gene is subcloned in the final plasmid vector which is selected from the group of pET vectors (Novagen). These vectors contain a strong T7 promoter. Preferably the plasmid vector pET3a is used. The expression plasmid which is thereby constructed is then transformed into the production strain which is selected from the group of strains which carry a chromosomal record for T7 RNA polymerase. Most preferably, *E. coli* BL21 (DE3) is used.

The procedure is continued with the preparation of inoculum and with the fermentation process. The fermentation can be performed at 37^0C, but is preferably performed at 25^0C.

The accumulated heterologous hG-CSF is found in the inclusion bodies and is suitable for the renaturation process and use in the isolation procedures.

<u>Description of the drawings:</u>

Figure 1: The scheme of the optimization steps of gene coding for hG-CSF

Figure 2: a) DNA sequence of the native gene coding for hG-CSF (GenBank: NM_000759)

b) DNA sequence of the optimized (Fopt5) gene coding for hG-CSF. The bases which differ from native gene are bolded.

Figure 3: a) SDS-PAGE (4 % stacking, 15 % separating; stained with Coomassie brilliant blue) of the samples of the proteins from the induced and noninduced cultures of production strains *E. coli* BL21 (DE3) with the

expression plasmid pET3a at 25° C and 42° C. The cultures were cultivated in the LBG10/amp100 medium.

Legend:

Load 1: BL21 (DE3) pET3a-hG-CSF non-induced at 25^{0}C (10 µl) (no traces of hG-CSF)

Load 2: BL21(DE3) pET3a-hG-CSF induced with IPTG at 25^{0}C (10 µl) (slight tracehG-CSF)

Load 3: BL21 (DE3) pET3a-hG-CSF non-induced at 42^{0}C (10 µl) (no traces hG-CSF)

Load 4: BL21 (DE3) pET3a-hG-CSF induced with IPTG at 42^{0}C (10 µl) (under 1 % hG-CSF)

Load 5: standard filgrastim 0.3 µg for Coomassie brilliant blue

Load 6: BL21 (DE3) pET3a-Fopt5 non-induced at 25^{0}C (5 µl) (6 % hG-CSF)

Load 7: BL21 (DE3) pET3a-Fopt5 induced with IPTG at 25^{0}C (5 µl) (over 50% hG-CSF)

b) detection with antibodies (Western blot); primary rabbit antibodies; secondary goat anti-rabbit IgG antibodies conjugated with horseradish peroxidase, substrate β-naphthol

The samples for the detection with antibodies were loaded in the same amount and in the same sequence as at SDS-PAGE (Figure 3a) with the exception of the standard which load was 0.08 µg.

Figure 4: SDS-PAGE (4 % stacking, 15 % separating; stained with Coomassie brilliant blue) samples of proteins from induced culture of the production strain *E. coli* BL21 (DE3) with the expression plasmid pET3a at 25° C. The cultures were cultivated in GYSP/amp100 and LYSP/amp100 medium.

Legend:

Load 1: LMW (BioRad)

Load 2: BL21 (DE3) pET3a/P-Fopt5, the culture cultivated in LYSP/amp100; (60% hG-CSF)

Load 3: BL21 (DE3) pET3a/P-Fopt5, the culture cultivated in LYSP/amp100; (over 54% hG-CSF)

Load 4: rhG-CSF (0.6 μg)

Load 5: rhG-CSF (1.5 μg)

Load 6: BL21 (DE3) pET3a/P-Fopt5, the culture cultivated in GYSP/amp100 (4 μl); (55% hG-CSF)

Load 7: BL21 (DE3) pET3a/P-Fopt5, the culture cultivated in GYSP/amp100 (5μl); (52% hG-CSF)

Examples:

Example 1: Construction of the optimal gene: Fopt5

Example 1a: The initial gene and plasmid preparations

The gene coding for hG-CSF was amplified from BBG13 (R&D) with the PCR method, which was also used to introduce by using the start oligonucleotides the restriction sites NdeI in BamHI at the start and terminal end of the gene. The gene was then incorporated in the plasmid pCytexΔH,H (see the description below) between the restriction sites NdeI in BamHI. All other optimization steps for the expression of the gene in *E. coli* were also performed in this plasmid.

During the initial gene preparation the EcoRV site was deleted (oligo M20z108) from the gene. This was performed with the aim to ensure the possibility of introduction of (individual) mutations by using the oligonucleotide-directed mutagenesis in the plasmid pCytexΔH,H with the kit Transformer™ Site-Directed Mutagenesis Kit (Clontech). The selection of mutants in the plasmid pCytexΔH,H-G-CSF via the restriction sites EcoRI/EcoRV was therefore possible.

The starting plasmid pCYTEXP1 (Medac, Hamburg) was reconstructed in a way to enable the constitutive expression. This was performed by the excision of the part of the gene coding for cI857 repressor between both restriction sites HindIII. The obtained plasmid was named pCytexΔH,H.

The oligonucleotide for the deletion of EcoRV site from the gene coding for hG-CSF:

M20z108 5' -CCT GGA AGG AAT ATC CCC CG-3'

Example 1b: Codon optimization (Figure 1)

In the first optimization step the synthetic gene between the restriction sites NdeI in SacI was constructed by ligation of five cassettes (A, B, C, D, E) which were composed of complementary oligonucleotides. This synthetic part of the gene represents the segment I. With the segment I the part of the native gene for hG-CSF gene between the restriction sites NdeI in SacI was replaced. This was performed by the excision of the first part of the gene between the restriction sites Nde I and SacI and its replacement with the synthetically prepared cassette. The process was performed in two steps. In the first step, the cassette A was ligated to the NdeI site and the cassette E was ligated to the SacI site. After 16 hours at 16^{0}C the ligation mixture was precipitated with ethanol to remove the excess of (not bound) oligonucleotides. In the second steps the central part of the whole cassette (cassette B, C and D) from the three previously ligated complementary nucleotides was added and the ligation was performed for 16 hours at 16^{0}C.

In the second optimization step the two t for *E. coli* most critical codons located in the segment III, namely, Arg148 in Gly150, were replaced by using the oligonucleotide-directed mutagenesis (TransformerTM Site-Directed Mutagenesis Kit (Clontech)).

In the third optimization step the segment IV was constructed in a similar way as the segment I with the exception of intermediate ethanol precipitation. The segment IV represents the last part of the gene between the restrictions sites NheI and BamHI and is composed of two pairs of complementary oligonucleotides (cassettes F in G).

In the fourth step of optimization the rare codon coding for Ile96 was replaced (ATA→ATT) (segment II) by using the oligonucleotide-directed mutagenesis (TransformerTM Site-Directed Mutagenesis Kit (Clontech)) and the restriction site for Apa I (Gly101 GGT→GGG) was introduced at the 3' end of the segment II.

Apa I restriction site was then used in the fifth optimization step with the aim to replace the native gene between SacI and ApaI with the synthetic DNA (segment II). This synthetic DNA is composed of three pairs of complementary oligonucleotides (cassette H, I and J). This was performed similarly as in the first step with the later addition of the cassette I.

1st optimization step:

complementary pairs of oligonucleotides (Nde I – Sac I; segment I in Figure 1):

Cassette A: composed of complementary oligonucleotides zg1os1 in sp1os2:

zg1os1 5' TAT GAC ACC ACT GGG TCC AGC TTC TTC TCT GCC GCA AAG 3'

sp1os2 5' GCA GAG AAG AAG CTG GAC CCA GTG GTG TCA 3'

Cassette B: composed of complementary oligonucleotides zg2os3 in sp2os4:

zg2os3 5' CTT TCT GTT GAA ATG TTT AGA ACA AGTTCG TAA AAT TCA AG 3'

sp2os4 5' GAA CTT GTT CTA AAC ATT TCA ACA GAA AGC TTT GCG 3'

Cassette C: composed of complementary oligonucleotides zg3os5 in sp3os6:

zg3os5 5' GTG ATG GTG CAG CTT TAC AAG AAA AAC TGT GTG 3'

sp3os6 5' GTT TTT CTT GTA AAG CTG CAC CAT CAC CTT GAA TTT TAC 3'

Cassette D: composed of complementary oligonucleotides zg4os7 in sp4os8:

zg4os7 5' CAA CTT ATA AAC TGT GTC ATC CAG AAG AAC TGG TTC TGT TAG 3'

sp4os8 5' CAG TTC TTC TGG ATG ACA CAG TTT ATA AGT TGC ACA CA 3'

Cassette E: composed of complementary oligonucleotides zg5os9 in sp5os10:

zg5os9 5' GTC ATT CTC TGG GTA TTC CGT GGG CTC CTC TGA GCT 3'

sp5os10 5' CAG AGG AGC CCA CGG AAT ACC CAG AGA ATG ACC TAA CAG AAC 3'

2nd optimization step: oligonucleotides for the replacement of the most critical codons by using the oligonucleotide-directed mutagenesis

replacement Arg 148 (CGG – CGT) in Gly 150 (GGA – GGT)

m38os16

5' CTC TGC TTT CCA GCG CCG TGC AGG TGG GGT CCT GGT TG 3'

3rd optimization step: complementary pairs of nucleotides (Nhe I – BamH I; segment IV on Figure 1):

Cassette F: composed of complementary nucleotides zg6os11 in sp6os12:

zg6os11 5' CTA GCC ATC TGC AAT CTT TTC TGG AAG TTA G 3'

sp6os12 5' ACG ATA GCT AAC TTC CAG AAA AGA TTG CAG ATG G 3'

Cassette G: composed of complementary oligonucleotides zg7os13 in sp7os14:

zg7os13 5' CTA TCG TGT TCT GCG TCA TCT GGC TCA GCC GTG ATA AG 3'

sp7os14 5' GAT CCT TAT CAC GGC TGA GCC AGA TGA CGC AGA AC 3'

4th optimization step: oligonucleotides for the introduction of Apa I (Gly101 GGT →GGG), and the replacement of the rare codon Ile96 by using the oligonucleotide-directed mutagenesis

insertion of Apa I (Gly101 GGT →GGG), and replacement Ile 96 (ATA – ATT):

Apalos15

5' GCC CTG GAG GGG ATT TCC CCC GAG TTG GGG CCC ACC TTG GAC AC 3'

5. optimization step: complementary pairs of oligonucleotides (Sac I – Apa I; segment II in Figure 1):

Cassette H: composed of complementary oligonucleotides zg8os18 in sp8os19:

zg8os18 5' CCT GTC CGA GCC AGG CGC TGC AGC TGG CAG GCT GCC TGA G 3'

sp8os19 5' CCT GCC AGC TGC AGC GCC TGG CTC GGA CAG GAG CT 3'

Cassette I: composed of complementary oligonucleotides zg9os20 in sp9os21:

zg9os20 5' CCA ACT GCA TAG CGG TCT GTT TCT GTA TCA GGG TCT GCT G 3'

sp9os21 5' CTG ATA CAG AAA CAG ACC GCT ATG CAG TTG GCT CAG GCA G 3'

Cassette J: composed of complementary oligonucleotides zg10os22 in sp10os23:

zg10os22 5' CAG GCG CTG GAA GGC ATT TCC CCG GAA CTG GGG CC 3'

sp10os23 5' CCA GTT CCG GGG AAA TGC CTT CCA GCG CCT GCA GCA GAC C 3'

Example 2: Expression of the synthetic gene coding for hG-CSF in *E. coli*

The optimized gene Fopt5 was excised from the plasmid pCyΔH,H with the restriction enzymes NdeI in BamHI and the gene was then subcloned in the final expression plasmid pET3a (Novagen, Madison USA) which was then transformed into the production strain *E. coli* BL21 (DE3).

The cultures were prepared on a shaker for 24 hours at 160 rpm at 25^{0}C or 15 hours at 42^{0}C:

- in LBG10/amp100 medium (10 g/l tryptone, 5g/l yeast extract, 10 g/l NaCl, 10 g/l glucose, 100 mg/L ampicillin). The induction was performed with the addition of IPTG to the final concentration of 0.4 mM.

The cultures were prepared on a shaker for 24 hours at 160 rpm at 25^{0}C:

- in GYSP/amp100 medium (20 g/l phytone, 5g/l yeast extract, 10 g/l NaCl, 10 g/l glucose, metals in traces, 100 mg/L ampicillin). The induction was performed with the addition of IPTG into the medium to the final concentration of 0.4 mM.
- in LYSP/amp100 medium (20 g/l phytone, 5g/l yeast extract, 10 g/l NaCl, 6 g/l glycerol, 4 g/l lactose, metals in traces, 100 mg/L ampicillin). The induction was performed with the addition of lactose into the medium.

The inoculum was prepared in LBG/amp100 medium (10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl, 2.5 g/L glucose) and 100 mg/L ampicillin at 25^{0}C, 160 rpm overnight.

For analysis 8 ml of the culture was centrifuged at 5000 rpm. The pellets were then resuspended in 10 mM TrisHCl/pH=8.0 in a proportion of 0.66 ml buffer added to calculated 1 unit OD_{600nm}. The loaded amounts were thereby equalized. Namely, the final OD_{600nm} of the cultures in the stated examples were not equal. The samples were mixed in the proportion of 3:1 with 4x SDS – sample buffer with DTT (pH=8.7) and heated 10 minutes at 95^{0}C, centrifuged and loaded onto he gel.

The proportions of accumulated hG-CSF found in the form of inclusion bodies for the native and optimized gene are described in Table 1.

Table 1: The comparison of the accumulation levels of hG-CSF for the native and the optimized gene (Fopt5)

		proportion (%) hG-CSF to total proteins		
Expression system	cultivation and induction conditions	native gene coding for hG-CSF		optimized gene Fopt5
	cultivation temperature	25° C	42° C	25° C
pET3a / *E. coli* BL21 (DE3)	medium LBG10/amp100 0.4 mM IPTG	traces	< 1 %	> 40 %
pET3a / *E. coli* BL21 (DE3)	medium GYSP/amp100	< 1 %	< 1 %	> 52 %
pET3a / *E. coli* BL21 (DE3)	medium LYSP/amp100	< 1 %	< 1 %	> 52 %

The indicated values for hG-CSF contents are obtained by the densitometric analysis of SDS-PAGE gels stained with Coomassie brilliant blue in the case of Fopt5 (Figure 3a in Figure 4) and by using the detection with antibodies in the case of unoptimized gene (Figure 3b). In the case of Fopt5 the relative proportion by the estimation of expression was determined with the profile analysis (program Molecular analyst; BioRad) of the gels by using the apparatus Imaging densitometer Model GS670 (BioRad).

Description of DNA sequences

<110> Lek Pharmaceuticals d. d.

<120> Synthetic gene coding for human granulocyte-colony stimulating factor for the expression in *E. coli*

<160> 2

<210> SEQ ID NO: 1

<211> 525 base pairs

<212> DNA

<213> synthetic sequence

<220> gene

<400> SEQ ID NO: 1

```
atgacaccac tgggtccagc ttcttctctg ccgcaaagct ttctgttgaa atgtttagaa    60
caagttcgta aaattcaagg tgatggtgca gctttacaag aaaaactgtg tgcaacttat   120
aaactgtgtc atccagaaga actggttctg ttaggtcatt ctctgggtat tccgtgggct   180
cctctgagct cctgtccgag ccaggcgctg cagctggcag gctgcctgag ccaactgcat   240
agcggtctgt ttctgtatca gggtctgctg caggcgctgg aaggcatttc cccggaactg   300
gggcccacct tggacacact gcagctggac gtcgccgact ttgccaccac catctggcag   360
cagatggaag aactgggaat ggcccctgcc ctgcagccca cccagggtgc catgccggcc   420
ttcgcctctg ctttccagcg ccgtgcaggt ggggtcctgg ttgctagcca tctgcaatct   480
tttctggaag ttagctatcg tgttctgcgt catctggctc agccg                   525
```

<210> SEQ ID NO: 2

<211> 528 base pairs

<212> DNA

<213> synthetic sequence

<220> gene

<400> SEQ ID NO: 2

```
atgacaccac tgggtccagc ttcttctctg ccgcaaagct ttctgttgaa atgtttagaa    60
caagttcgta aaattcaagg tgatggtgca gctttacaag aaaaactgtg tgcaacttat   120
aaactgtgtc atccagaaga actggttctg ttaggtcatt ctctgggtat tccgtgggct   180
cctctgagct cctgtccgag ccaggcgctg cagctggcag gctgcctgag ccaactgcat   240
agcggtctgt ttctgtatca gggtctgctg caggcgctgg aaggcatttc cccggaactg   300
gggcccacct tggacacact gcagctggac gtcgccgact ttgccaccac catctggcag   360
cagatggaag aactgggaat ggcccctgcc ctgcagccca cccagggtgc catgccggcc   420
ttcgcctctg ctttccagcg ccgtgcaggt ggggtcctgg ttgctagcca tctgcaatct   480
tttctggaag ttagctatcg tgttctgcgt catctggctc agccgtga                528
```

Lek Pharmaceuticals d.d.

Patent claims

1. A DNA sequence coding for hG-CSF characterized in that the sequence comprises the nucleotide sequence of SEQ ID NO: 1.
2. A DNA sequence characterized in that the sequence comprises a nucleotide sequence selected from the group comprising a partial sequence of SEQ ID NO: 1 and nucleic acids which hybridize with the sequence of SEQ ID NO: 1 under stringent conditions.
3. An expression plasmid characterized in that the plasmid comprises the DNA sequence according to claim 1 and a plasmid vector
4. An expression plasmid characterized in that the plasmid comprises a DNA sequence according to claim 2 and a plasmid vector.
5. The expression plasmid according to claims 2 and 4 characterized in that the plasmid vector is selected from the group of pET vectors.
6. An expression system for the expression of DNA sequence according to claim 1 characterized in that the system comprises the DNA sequence, a plasmid vector and a production strain *E. coli*.
7. An expression system for the expression of the DNA sequence according to claim 2 characterized in that the system comprises the DNA sequence, a plasmid vector and a production strain *E. coli*.
8. The expression system according to claims 6 and 7 characterized in that the plasmid vector is selected from the group of pET vectors.
9. The expression system according to claims 6 in 7 characterized in that the production strain is *E. coli* BL21 (DE3).
10. A process for construction of DNA sequence according to claim 1 characterized in that the process comprises the methods selected from the group comprising:
    -replacement of some *E. coli* rare codons with *E. coli* preference codons,
    -replacement of some GC rich regions with AT rich regions
    and further comprises a completely unchanged part of the native sequence coding for hG-CSF.
11. A process for construction of DNA sequence according to claim 10 characterized in that the process does not involve changes in the regions from the group

comprising: translation initiation region, ribosome binding site and the region between the start codon and the ribosome binding site.

12. Expression of DNA sequence according to claim 1 in *E. coli*.
13. Expression of DNA sequence according to claim 2 in *E. coli*.

Lek Pharmaceuticals d.d.

Abstract

The invention relates to the synthetic gene coding for hG-CSF which enables expression in *E. coli* with the expression level being more than 52% of the recombinant hG-CSF regarding the total cellular proteins after expression.

Figure 1

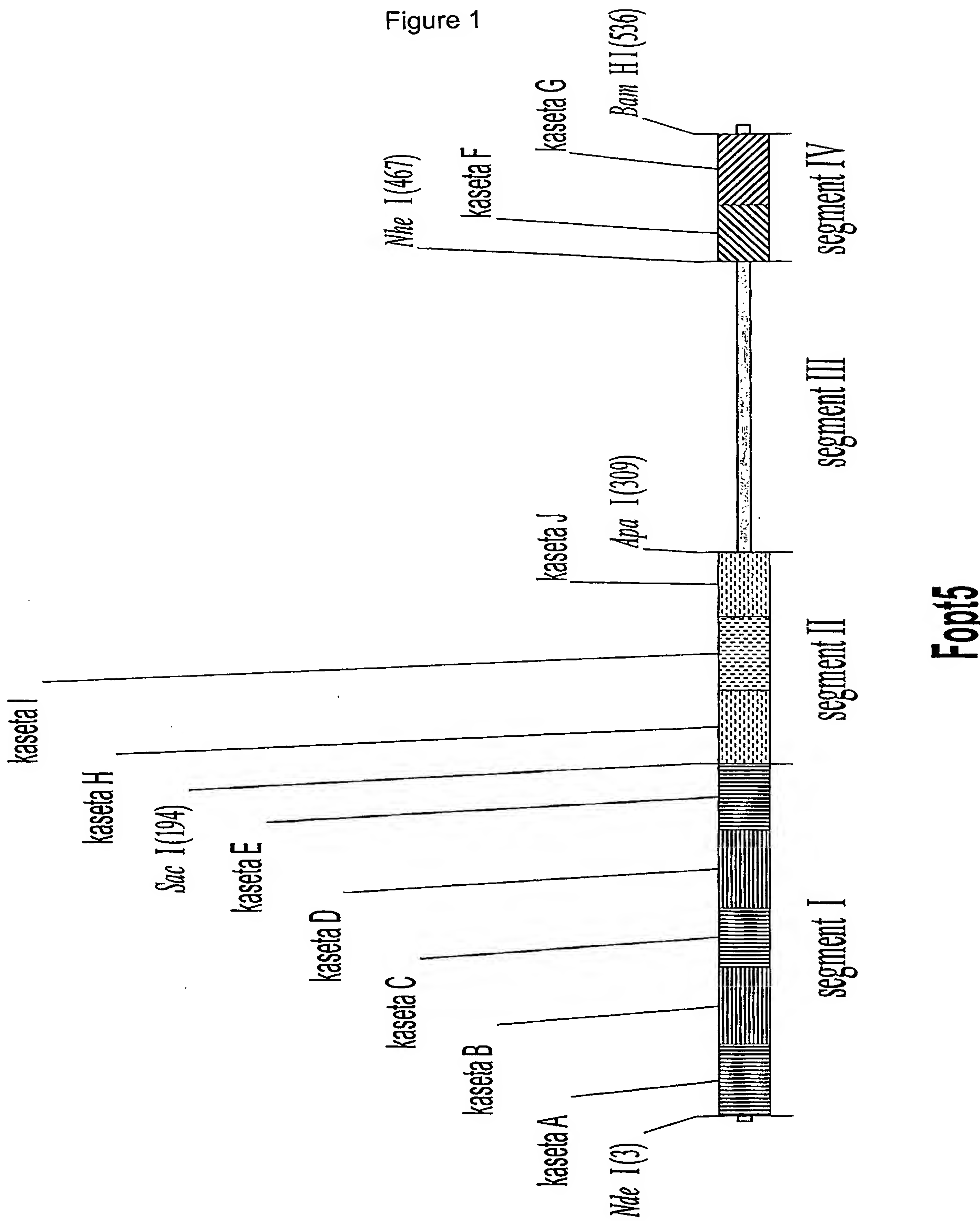

Nde I(3)
kaseta A
kaseta B
kaseta C
kaseta D
kaseta E
Sac I(194)
kaseta H
kaseta I
kaseta J
Apa I(309)
Nhe I(467)
kaseta F
kaseta G
Bam HI(536)
segment I
segment II
segment III
segment IV
Fopt5

Figure 2

a)

ATGACCCCCCTGGGCCCTGCCAGCTCCCTGCCCCAGAGCTTCCTGCTCAAGTG
CTTAGAGCAAGTGAGGAAGATCCAGGGCGATGGCGCAGCGCTCCAGGAGAAGC
TGTGTGCCACCTACAAGCTGTGCCACCCCGAGGAGCTGGTGCTGCTCGGACAC
TCTCTGGGCATCCCCTGGGCTCCCCTGAGCTCCTGCCCCAGCCAGGCCCTGCA
GCTGGCAGGCTGCTTGAGCCAACTCCATAGCGGCCTTTTCCTCTACCAGGGGC
TCCTGCAGGCCCTGGAAGGATATCCCCGAGTTGGGTCCCACCTTGGACACA
CTGCAGCTGGACGTCGCCGACTTTGCCACCACCATCTGGCAGCAGATGGAAGA
ACTGGGAATGGCCCCTGCCCTGCAGCCCACCCAGGGTGCCATGCCGGCCTTCG
CCTCTGCTTTCCAGCGCCGGGCAGGAGGGGTCCTGGTTGCTAGCCATCTGCAG
AGCTTCCTGGAGGTGTCGTACCGCGTTCTACGCCACCTTGCGCAGCCC

b)

ATGAC**A**CC**A**CTGGG**T**CC**A**GC**TTCT**TC**T**CTGCC**G**CA**A**AGCTT**T**CTG**T**T**G**AA**A**TG
TTTAGA**A**CAAGT**TC**G**T**AA**A**AT**T**CA**A**GG**T**GATGG**T**GCAGC**TT**T**A**CA**A**GA**A**AA**A**C
TGTGTGC**A**AC**T**TA**T**AA**A**CTGTG**T**CA**T**CC**A**GA**A**GA**A**CTGGT**T**CTG**T**T**A**GG**T**CA**T**
TCTCTGGG**T**AT**T**CC**G**TGGGCTCC**T**CTGAGCTCCTG**T**CC**G**AGCCAGGC**G**CTGCA
GCTGGCAGGCTGC**C**TGAGCCAACT**G**CATAGCGG**T**CT**G**TT**T**CT**G**TA**T**CAGGG**T**C
T**G**CTGCAGGC**G**CTGGAAGG**C**AT**T**TCCCC**G**GA**AC**TGGG**G**CCCACCTTGGACACA
CTGCAGCTGGACGTCGCCGACTTTGCCACCACCATCTGGCAGCAGATGGAAGA
ACTGGGAATGGCCCCTGCCCTGCAGCCCACCCAGGGTGCCATGCCGGCCTTCG
CCTCTGCTTTCCAGCGCCG**T**GCAGG**T**GGGGTCCTGGTTGCTAGCCATCTGCA**A**
TCTTT**T**CTGGA**A**GT**TAGC**TA**T**CG**T**GTTCT**G**CG**T**CA**T**CT**G**GC**T**CAGCC**G**

Figure 3

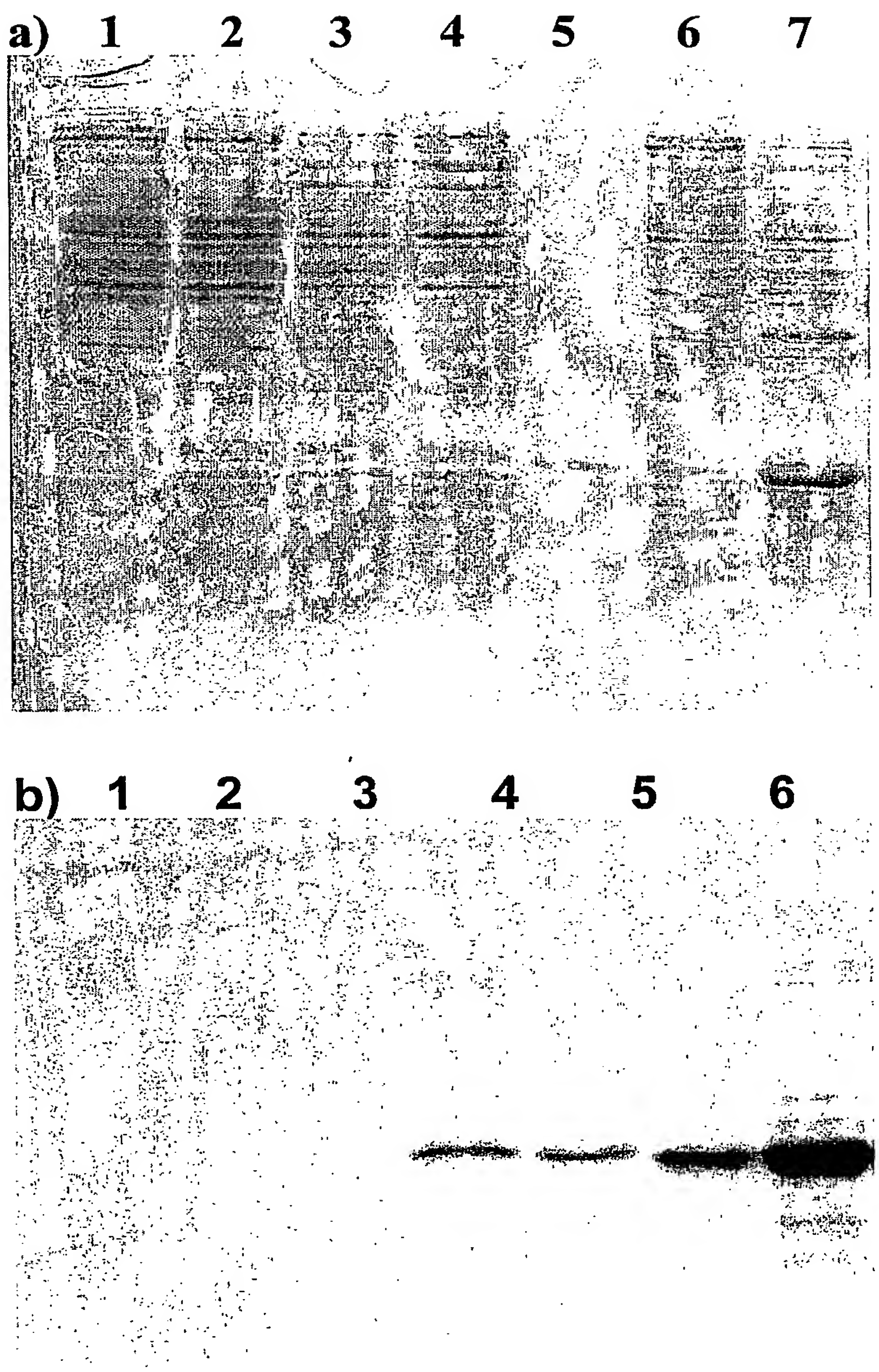
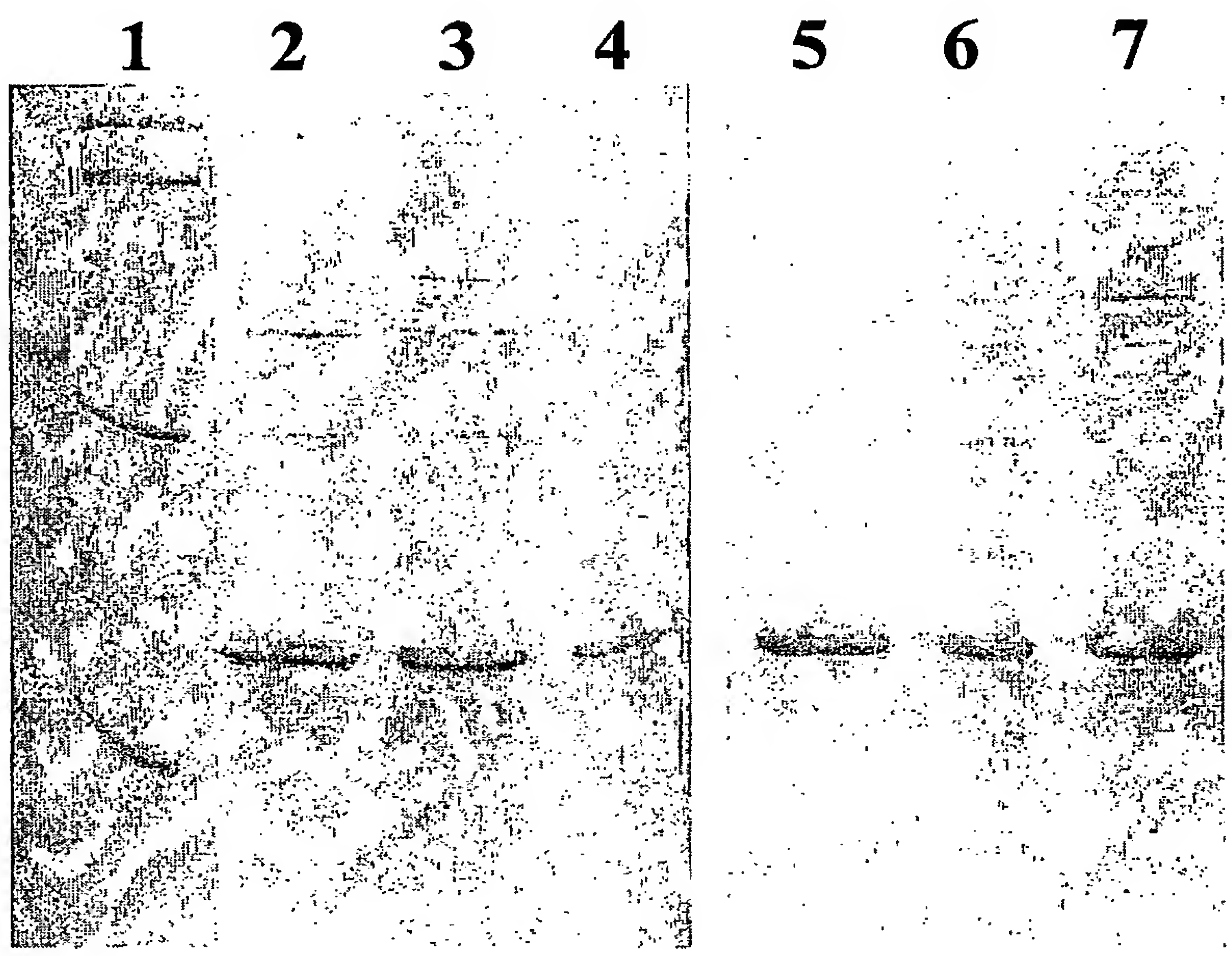


Figure 4



Lek Pharmaceuticals d.d.

The undersigned Djurdjica Mandrino, permanent court interpreter for the English language, appointed by Decree No. 756-4/91, issued on 11th of February 1991 by the Ministry of Justice and Administration, Republic of Slovenia, hereby declares that this document entirely corresponds to the original Slovene text.

Ljubljana, 12th June 2003

STALNI SODNI TOLMAČ ZA ANGLEŠKI IN SRBOHRVATSKI JEZIK
MANDRINO
DJURDJICA
Ane Ziherlove 4/29
LJUBLJANA